Observations on Cooperation

Yuval Heller and Erik Mohlin

Bar Ilan Univeristy, Lund University

2 February 2017

Observations on Cooperation

Yuval Heller* and Erik Mohlin‡

February 2, 2017

Abstract

This paper develops a new theory of community enforcement that explains how cooperation can be sustained when agents change their partners over time. We study environments in which agents are randomly matched to play a Prisoner's Dilemma, and each player observes a few of the partner's past actions against previous opponents. We depart from the existing related literature by allowing a small fraction of the population to be commitment types. The presence of committed agents destabilizes all previously proposed mechanisms for sustaining cooperation (e.g., contagious equilibria and belief-free equilibria). We present a novel, yet intuitive, combination of strategies that sustains cooperation in various environments. This mechanism is fully decentralized in the sense that each player's strategy conditions on only a few observations that the player makes regarding her current partner's past behavior. Moreover, we show that under an additional assumption of stationarity, this combination of strategies is essentially the *unique* mechanism to support full cooperation, and it is robust to various perturbations. Finally, we extend the results to a setup in which agents also observe actions played by past opponents against the current partner, and we characterize which observation structure is optimal for sustaining cooperation.

JEL Classification: C72, C73, D83. **Keywords**: Community enforcement; indirect reciprocity; random matching; Prisoner's Dilemma; image scoring.

1 Introduction

Consider the following example of a simple yet fundamental economic interaction. Alice has to trade with another agent, Bob, whom she does not know. Both sides have opportunities to cheat, to their own benefit, at the expense of the other. Alice is unlikely to interact with Bob again, and thus her ability to retaliate, in case Bob acts opportunistically, is restricted. The effectiveness of external enforcement is also limited, e.g., due to incompleteness of contracts, non-verifiability of information, and court costs. Thus cooperation may be impossible to achieve. Alice searches for information about Bob's past behavior, and she obtains anecdotal

*Affiliation: Department of Economics, Bar Ilan University, Israel. E-mail: yuval.heller@biu.ac.il.

†Affiliation: Department of Economics, Lund University, Sweden. E-mail: erik.mohlin@nek.lu.se.

‡A previous version of this paper was circulated under the title "Stable observable behavior." We have benefited greatly from discussions with Vince Crawford, Eddie Dekel, Christoph Kuzmics, Ariel Rubinstein, Larry Samuelson, Bill Sandholm, Rann Smorodinsky, Rani Spiegler, Balázs Szentes, Satoru Takahashi, Jörgen Weibull, and Peyton Young. We would like to express our deep gratitude to seminar/workshop participants at the University of Amsterdam (CREED), University of Bamberg, Bar Ilan University, Bielefeld University, University of Cambridge, Hebrew University of Jerusalem, Helsinki Center for Economic Research, Interdisciplinary Center Herzliya, Israel Institute of Technology, Lund University, University of Oxford, University of Pittsburgh, Stockholm School of Economics, Tel Aviv University, NBER Theory Workshop at Wisconsin-Madison, KAEA session at the ASSA 2015, and the Biological Basis of Preference conference at Simon Fraser University, for many useful comments. Yuval Heller is grateful to the European Research Council for its financial support (starting grant #677057). Erik Mohlin is grateful to Handelsbankens forskningsstiftelser (grant #P2016-0079:1) for its financial support. Last but not least, we thank Renana Heller for suggesting the title.

evidence about Bob's actions in a couple of past interactions. Alice considers this information when she decides how to act. Alice also takes into account that her behavior toward Bob in the current interaction may be observed by her future partners. Historically, the above-described situation was a challenge to the establishment of long-distance trade (Milgrom, North, and Weingast, 1990; Greif, 1993), and it continues to play an important role in the modern economy, in both offline (Bernstein, 1992; Dixit, 2003) and online interactions (Resnick and Zeckhauser, 2002; Jøsang, Ismail, and Boyd, 2007).

Several papers have studied the question of how cooperation can be supported by means of community enforcement. We differ from the existing community enforcement literature in that we allow a few agents in the population to be committed to behaviors that do not necessarily maximize their payoffs. It turns out that this seemingly small modification completely destabilizes existing mechanisms for sustaining cooperation when agents are randomly matched with new partners in each period. Specifically, both the contagious equilibria (Kandori, 1992; Ellison, 1994)[1] and the "belief-free" equilibria (Takahashi, 2010; Deb, 2012)[2] fail in the presence of a small fraction of committed agents.

Our key results are as follows.[3] First, we show that always defecting is the unique perfect equilibrium, regardless of the number of observed actions, provided that the bonus of defection in the underlying Prisoner's Dilemma is larger when the partner cooperates than when the partner defects. Second, in the opposite case, when the bonus of defection is larger when the partner defects than when the partner cooperates, we present a novel and essentially unique combination of strategies that sustains cooperation: all agents cooperate when they observe no defections and defect when they observe at least two defections. Some of the agents also defect when observing a single defection. Importantly, this cooperative behavior is robust to various perturbations, and it appears consistent with experimental data. Third, we extend the model to environments in which an agent also obtains information about the behavior of past opponents against the current partner. We show that in this setup cooperation can be sustained if and only if the bonus of defection of a player is less than half the loss she induces to a cooperative partner. Finally, we characterize an observation structure that allows cooperation to be supported as a perfect equilibrium action in *all* Prisoner's Dilemma games. In all observation structures we use the same essentially unique construction to sustain cooperation.

We note that the effect of commitment types in our setup is substantially different from the effect of commitment types in the existing literature on reputation in repeated games, in which a stylized main result is that there exists a relatively high lower bound on all equilibrium payoffs. By contrast, the introduction of commitment types in our model (1) uniquely selects the equilibrium with the lowest payoff in some environments, and (2) implies a unique way to support the efficient cooperative equilibrium in other environments (see the more thorough discussion in Section 2.6).

Overview of the Model Agents in an infinite population are randomly matched into pairs to play a symmetric one-shot game. Before playing the game, each agent privately draws a random sample of k actions that

[1] In contagious equilibria players start by cooperating. If one player defects at stage t, her partner defects at stage $t+1$, infecting another player who defects at stage $t+2$, and so on. The non-robustness of these equilibria to a single "crazy" agent was already noted by Ellison (1994, p. 578): "If one player were 'crazy' and always played D (or simply was unaware which equilibrium was being played) again the contagious strategies would not support cooperation. In large populations, the assumption that all players are rational and know their opponents' strategies may be both very important to the conclusions and fairly implausible."

[2] In belief-free equilibria players are always indifferent between their actions, but they choose different mixed actions depending on the signal they obtain about the partner. To the best of our knowledge we are the first to show the non-robustness of these equilibria to the presence of a few committed agents. Elsewhere, one of us develops a somewhat related critique on belief-free equilibria in a standard setup of repeated games between the same two players (Heller, 2017).

[3] As discussed later, our uniqueness results also rely on an additional assumption that agents are restricted to choose stationary strategies, which depend only on the signal about the partner. As shown in Section 6, all other results hold also in a standard setup without the restriction to stationary strategies.

have been played by her partner against other opponents in the past.[4] The assumption that a small random sample is taken from the entire history of the partner is intended to reflect a setting in which the memory of past interactions is long and accurate but dispersed. This means that the information that reaches an agent about her partner (through gossip) arrives in a non-deterministic fashion and may stem from any point in the past.

We require each agent to follow a *stationary strategy*, i.e., a mapping that assigns a mixed action to each signal that the agent may observe about the current partner. (That is, the action is not allowed to depend on calendar time or on the agent's own history.)[5] A *steady state* of the environment is a pair consisting of: (1) a distribution of strategies with a finite support that describes the fractions of the population following the different strategies, and (2) a *signal profile* that describes the distribution of signals that is observed when an agent is matched with a partner playing any of the strategies present in the population. The signal profile is required to be *consistent* with the distribution of strategies in the sense that a population of agents who follow the distribution of strategies and observe signals about the partners sampled from the signal profile will behave in a way that induces the same signal profile.[6]

Our restriction to stationary strategies and our focus on consistent steady states allows us to relax the standard assumption that there is an initial time zero at which the entire community starts to interact (see, e.g., Kandori, 1992; Dixit, 2003; Deb and González-Díaz, 2014). In many real-life situations, the interactions within a community have been going on from time immemorial. Consequently the participants may have only a vague idea of the starting point. It seems implausible that agents would be able to condition their behavior on everything that has happened since then (or on "calendar time"). A detailed discussion of this issue and its relation to the existing literature appears in Section 2.6.

We perturb the environment by introducing ϵ *committed agents* who each follow one strategy from an arbitrary finite set of *commitment strategies*.[7] We assume that at least one of the commitment strategies is totally mixed, which implies that all signals (i.e., all sequences of k actions) are observed with positive probability. A *steady state* in a perturbed environment describes a population in which $1-\epsilon$ of the agents are *normal,* i.e., they play strategies that maximize their long-run payoffs, while ϵ of the agents follow commitment strategies.

We adapt the notions of Nash equilibrium, perfect equilibrium (Selten, 1975), and strict perfection (Okada, 1981) to our setup.[8] A steady state is a *Nash equilibrium* if no normal agent can gain in the long run by deviating to a different strategy.[9] The deviator's payoff is calculated in the new steady state that emerges following her deviation. A steady state is a *perfect equilibrium* if it is the limit of a sequence of Nash equilibria in a converging sequence of perturbed environments. A pure action a^* is a *strictly perfect equilibrium action* if, for *any* converging sequence of perturbed environments, there is a converging sequence of Nash equilibria such that in

[4] In the main model these k actions are sampled from the entire history of play of the partner. In Section 6, we present a variant of the model in which each agent observes the most recent k actions of the partner.

[5] This assumption is relaxed in Section 6.

[6] The reason why the consistent signal profile is required to be part of the description of a steady state, rather than being uniquely determined by the distribution of strategies, is that our environment, unlike a standard repeated game, lacks a global starting time that determines the initial conditions. An example of a strategy that has multiple consistent signal profiles is as follows. The underlying game is the Prisoner's Dilemma, k is equal to three, and everyone plays the most frequently observed action in the sample of the three observed actions. There are three behaviors that are consistent with this population: one in which everyone cooperates, one in which everyone defects, and one in which everyone plays (on average) uniformly. In Section 2.6 we discuss our modeling choice of not having a calendar time, and in Section 6 we show that most of our results hold also in a conventional model with a global starting time.

[7] In Section 7.3 we discuss how to extend our results to more general perturbed environments in which, in addition to commitment strategies, there are also observation errors or trembles.

[8] In Appendix B we show that our perfect equilibria also satisfy the refinement of evolutionary stability (Maynard Smith, 1974).

[9] In the main model players are assumed to be arbitrarily patient. The alternative model of Section 6 relaxes this assumption and introduces a discount factor.

the limit everyone plays a^*. That is, strict perfection requires stability with respect to *all* commitment strategies, whereas the stability of a perfect equilibrium may rely on the absence of some commitment strategies.[10]

Summary of Results We begin our analysis with two results for general games. Our first result is that any Nash equilibrium of the underlying game can be implemented as a Nash equilibrium of the environment, for any value of k. Similarly, any perfect equilibrium of the underlying game can be implemented as a perfect equilibrium of the environment. Next, we demonstrate the usefulness of the refinement of strict perfection by showing that in coordination games only the Pareto-efficient Nash equilibrium satisfies strict perfection whenever agents observe at least two actions.

The remaining results of the paper focus on the Prisoner's Dilemma game, in which each player decides simultaneously whether to cooperate or defect (see the payoff matrix in Table 1); if both players cooperate they obtain a payoff of one, if both defect they obtain a payoff of zero, and if one of the players defects, the defector gets $1+g$, while the cooperator gets $-l$, where $g, l > 0$ and $g < l + 1$. (The latter inequality implies that mutual cooperation is the efficient outcome that maximizes the sum of payoffs.) We say that a Prisoner's Dilemma game is *offensive* if there is a stronger incentive to defect against a cooperator than against a defector (i.e., $g > l$); in a *defensive* Prisoner's Dilemma the opposite holds[11] (i.e., $g < l$). We start with a simple result (Prop. 4) that shows that defection is a strictly perfect equilibrium action for any number of observed actions.

Table 1: Matrix Payoffs of Prisoner's Dilemma Games

	c	d
c	1 , 1	$-l$, $1+g$
d	$1+g$, $-l$	0 , 0

$g, l > 0$, $g < l + 1$

Our first main result (Theorem 1) shows that always defecting is the unique perfect equilibrium in any offensive Prisoner's Dilemma game (i.e., $g > l$) for any number of observed actions. The result assumes a mild *regularity* condition on the set of commitment strategies (Def. 3), namely, that this set is rich enough that, in any steady state of the perturbed environment, at least one of the commitment strategies induces agents to defect with a different probability than some of the normal agents.[12] The intuition is as follows. The mild assumption that not all agents defect with exactly the same probability implies that the signal that Alice observes about her partner Bob is not completely uninformative. In particular, the more often Alice observes Bob to defect, the more likely Bob will defect against Alice. In offensive games, it is better to defect against partners who are likely to cooperate than to defect against partners who are likely to defect. This implies that

[10] The equilibria presented in our main results also satisfy a refinement of *robustness*, i.e., the condition that no small perturbation in the distribution of observed signals can move the population's behavior away from a situation in which everyone plays the equilibrium outcome (see Definition 8).

[11] This follows the terminology of Dixit (2003). Takahashi (2010) calls offensive (defensive) PDs submodular (supermodular). If cooperating is interpreted as exerting high effort, then the defensive Prisoner's Dilemma exhibits strategic complementarity: increasing one's effort from low to high is less costly if the opponent exerts high effort (as illustrated in Example 2 in Sect. 4.1).

[12] Propositions 1–2 study the implications of the mild regularity refinements in other games. Specifically, they show that (1) any perfect equilibrium of the underlying game that is not totally mixed can be implemented as a regular perfect equilibrium in any environment, and (2) the mild refinement rules out some totally mixed perfect equilibria of the underlying game, such as the totally mixed equilibrium in a coordination game.

a deviator who always defects is more likely to induce normal partners to cooperate. Consequently, such a deviator will outperform any agent who cooperates with positive probability.

Theorem 1 may come as a surprise in light of a number of existing papers that have presented various equilibrium constructions that support cooperation in any Prisoner's Dilemma game that is played in a population of randomly matched agents. As mentioned above (and discussed in more detail in Section 4.2), our result demonstrates that, in the presence of a small fraction of committed agents, the mechanisms that have been proposed to support cooperation fail, regardless of how these committed agents play. In this way our paper provides a theoretical explanation of why experimental evidence suggests that subjects' behavior corresponds neither to contagious equilibria (see, e.g., Duffy and Ochs, 2009) nor to belief-free equilibria (see, e.g., Matsushima, Tanaka, and Toyama, 2013).[13] *Additional empirical predictions of our model are discussed in Section 7.2.*

Our second main result (Theorem 2) shows that cooperation is a strictly perfect equilibrium action in any defensive Prisoner's Dilemma game ($g < l$) when players observe at least two actions. Moreover, there is an essentially unique distribution of strategies that support cooperation, according to which: (a) all agents cooperate when observing no defections, (b) all agents defect when observing at least 2 defections, (c) the normal agents defect with an average probability of $0 < q < 1$ when observing a single defection.[14] The intuition for the result is as follows. Defection yields a direct gain that is increasing in the partner's probability of defection (due to the game being defensive). In addition, defection results in an indirect loss because it induces future partners to defect when they observe the current defection. This indirect loss is independent of the current partner's behavior. One can show that there always exists a probability q such that the above distribution of strategies balances the direct gain and the indirect loss of defection, conditional on the agent observing a single defection. Furthermore, cooperation is the unique best reply conditional on the agent observing no defections, and defection is the unique best reply conditional on the agent observing at least two defections.

Next, we analyze the case of the observation of a single action (i.e., $k = 1$). Prop. 5 shows that cooperation is a perfect equilibrium action in a defensive Prisoner's Dilemma if and only if the bonus of defection is not too large (specifically, $g \leq 1$). The intuition is that similar arguments to the result above imply that there exists a unique average probability $q < 1$ by which agents defect when observing a defection in any cooperative perfect equilibrium. This implies that a deviator that always defects succeeds in getting a payoff of $1 + g$ in a fraction $1 - q > 0$ of the interactions, and that such a deviator outperforms the incumbents if g is too large.

Observations Based on Action Profiles So far we have assumed that each agent observes only the partner's (Bob's) behavior against other opponents, but that she cannot observe the behavior of the past opponents against Bob. In Section 5 we relax this assumption. Specifically, we study three observation structures: the first two seem to be empirically relevant, and the third one is theoretically important since it allows us to construct an equilibrium that sustains cooperation in all Prisoner's Dilemma games.

[13] Theorem 1 also shows that even when an agent observes many of her partner's actions, there is no way in which she can use this information to assess her partner's reputation, along the lines of the binary reputation mechanisms of Sugden (1986) and Kandori (1992, Theorem 2) (see also the related models in Okuno-Fujiwara and Postlewaite, 1995 and Ohtsuki and Iwasa, 2006). In these reputation models each agent starts with a "good label.. This label automatically becomes "bad" if a player defects against a "good" partner. The equilibrium strategy that supports full cooperation, given an exogenous mechanism that induces these labels, is to cooperate against "good" partners and defect against "bad" partners. Our result shows that despite their seemingly simple structure, reputation mechanisms cannot be implemented by plausible decentralized observation structures in the presence of a few committed agents.

[14] The specific commitment strategies that are present in the perturbed environment influence two aspects of the perfect equilibrium that supports cooperation: (1) they affect the average defection probability when an agent observes a single defection, and (2) they determine whether each agent mixes when she observes a single defection or whether the population is composed of two different groups of agents, such that only agents in one of these groups defect when they observe a single defection.

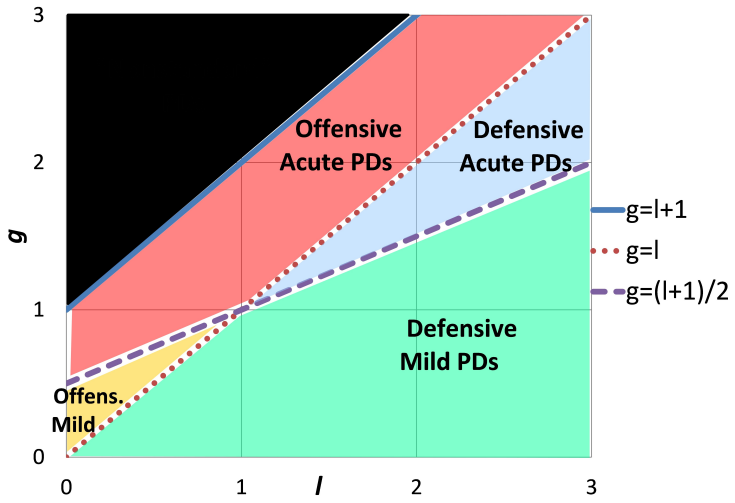
1. *Observing conflicts*: each agent observes, in each of the k sampled interactions of her partner, whether there was mutual cooperation (i.e., no conflict; both partners are "happy") or not (i.e., partners complain about each other, but it is too costly for an outside observer to verify who actually defected). Such an observation structure (which we have not seen described in the existing literature) seems like a plausible way to capture non-verifiable feedback about the partner's behavior.

2. *Observing action profiles:* each agent observes the full action profile in each of the sampled interactions.

3. *Observing actions against cooperation:* each agent observes in each of the sampled interactions, what action the partner took provided that the partner's opponent cooperated. If the partner's opponent defected then there is no information about what the partner did.

It turns out that the stability of cooperation in the first two observation structures crucially depends on a novel classification of Prisoner's Dilemma games. We say that a Prisoner's Dilemma game is *acute* if $g > \frac{l+1}{2}$, and *mild* if $g < \frac{l+1}{2}$. The threshold between the two categories, namely, $g = \frac{l+1}{2}$, is characterized by the fact that the gain from a single unilateral defection is exactly half the loss incurred by the partner who is the sole cooperator. Consider a setup in which an agent is deterred from unilaterally defecting because it induces future partners to unilaterally defect against the agent with some probability. Deterrence in acute Prisoner's Dilemmas requires this probability to be more than 50%, while a probability of below 50% is enough to deter deviations in mild PDs. Figure 1 illustrates the classification of games into mild/acute and offensive/defensive (see Section 5.2 for further discussion).

Figure 1: Classification of Prisoner's Dilemma Games



Our next results (Theorems 3–4) show that in both observation structures (conflicts or action profiles, and any $k \geq 2$) cooperation is a perfect equilibrium action if and only if the underlying Prisoner's Dilemma game is mild. Moreover, cooperation is supported by essentially the same unique behavior as in Theorem 2. The intuition for why cooperation cannot be sustained in acute games with observation of conflicts is as follows. In order to support cooperation agents should be deterred from defecting against cooperators. As discussed above, in acute games, such deterrence requires that each such defection induces future partners to defect with a probability of at least 50%. However, this requirement implies that defection is contagious: each defection by an

agent makes it possible that future partners observe a conflict both when being matched with the defecting agent, and when being matched with the defecting agent's partner. Such future partners defect with a probability of at least 50% when making such observations. Thus the fraction of defections grows steadily, until all normal agents defect with high probability.

The intuition for why cooperation cannot be sustained in acute games with observation of action profiles is as follows. The fact that deterring defections in acute games requires future partners to defect with a probability of at least 50% when observing a defection implies that when an agent (Alice) observes her partner (Bob) to defect against a cooperative opponent, then Bob is more likely to do so because he is a normal agent who observed his past opponent to defect than because Bob is a committed agent. This implies that Alice puts a higher probability on Bob defecting against her if she observes Bob to have defected against a partner who also defected than she does if she observes Bob to have defected against an opponent who cooperated. Thus, defecting is the unique best reply when observing the partner defect against a defector, but it removes the incentives required to support stable cooperation.

Finally, we show that the third observation structure, *observing actions against cooperation*, is optimal in the sense that it sustains cooperation as a perfect equilibrium action for any Prisoner's Dilemma game (Theorem 5). The intuition for this result is that not allowing Alice to observe Bob's behavior against a defector helps to sustain cooperation because it implies that defecting against a defector does not have any negative indirect effect (in any steady state) because it is never observed by future opponents. This encourages agents to defect against partners who are more likely to defect (regardless of the values of g and l).

Conventional Model and Unrestricted Strategies In Section 6, we relax the assumption that agents are restricted to choosing only stationary strategies. We present a conventional model of repeated games with random matching that differs from the existing literature only by introducing a few committed agents. We show that this difference is sufficient to yield most of our key results.[15]

Table 2: Summary of Key Results: When Is Cooperation a Perfect Equilibrium Outcome?

Category of PD	Parameters	Observation Structure (any $k \geq 2$)			
		Actions	Conflicts	Action profiles	Actions against cooperation
Mild & Defensive	$g < \min\left(l, \frac{l+1}{2}\right)$	**Y**	**Y**	**Y**	**Y**
Mild & Offensive	$l < g < \frac{l+1}{2}$	**N**			
Acute & Defensive	$\frac{l+1}{2} < g < l$	**Y**	**N**	**N**	
Acute & Offensive	$\max\left(l, \frac{l+1}{2}\right) < g$	**N**			

Specifically, the characterization of the conditions under which cooperation can be sustained as a perfect equilibrium outcome (as summarized in Table 1 below) holds also when agents are not restricted to stationary strategies, and even when agents observe the most recent past actions of the partner. On the other hand, the relaxation of the stationarity assumption in Section 6 weakens the uniqueness results of the main model in two respects: (1) rather than showing that defection is the unique equilibrium outcome in offensive games, we show only that it is impossible to sustain full cooperation in such games; and (2) while a variant of the simple strategy

[15]For brevity, we focus only on the formal results on the observations of actions (Theorem 6). The adaptation of the results on general observation structures is analogous.

of the main model still supports cooperation when the set of strategies is unrestricted, we are no longer able to show that this strategy is the unique way to support full cooperation.[16]

Structure Section 2 presents the model. Our solution concept is described in Section 3. Section 4 contains our main results. Section 5 extends the model to deal with general observation structures. Section 6 adapts our key result to a conventional model with an unrestricted set of strategies. In Section 7 we discuss the related literature, our empirical predictions, the robustness of our results, and directions for future research. Appendix A presents an example of a perfect equilibrium with partial cooperation. Appendix B presents the refinement of evolutionary stability. Appendix C studies the introduction of cheap talk to our setup. The formal proofs appear in Appendix D.

2 Stationary Model

2.1 Environment

We model an environment in which patient agents in a large population are randomly matched at each round to play a two-player symmetric one-shot game. For tractability we assume throughout the paper that the population is a continuum.[17] In the main model we further assume that the agents are infinitely lived and do not discount the future (i.e., they maximize the average per-round long-run payoff). Alternatively, our main model can be interpreted as representing interactions between finitely lived agents who belong to infinitely lived dynasties, such that an agent who dies is succeeded by a protégé who plays the same strategy as the deceased mentor, and each agent observes k random actions played by the partner's dynasty.

Before playing the game, each agent (she) privately observes k random actions that her partner (he) played against other opponents in the past.[18] As described in detail below, in the baseline model agents are restricted to using only stationary strategies, such that each agent's behavior depends only on the signal about the partner, and not on the agent's own past play or on time. Thus, if all agents observe signals that come from a stationary distribution then the agents' behavior will result in a well-defined aggregate distribution of actions that is also stationary. We focus on steady states of the population, in which the distribution of actions, and hence the distribution of signals, is indeed stationary. In such steady states, the k actions that an agent observes about her partner are drawn independently from the partner's stationary distribution of actions. This sampling procedure may be interpreted as the limit of a process in which each agent randomly observes k actions that are uniformly sampled from the last n interactions of the partner, as $n \to \infty$.

Remark 1. In Section 6 we relax some of the above-mentioned simplifying assumptions, and we study setups in which agents are finitely lived (and start with an empty history of actions), agents observe their partners' most recent actions, or agents are allowed to choose non-stationary strategies.

[16] In addition, in the setup of Section 6 we show only that cooperation is a perfect equilibrium outcome, rather than that it satisfies the refinement of strict perfection.

[17] The results can be adapted to a setup with a large finite population. We do not formalize a large finite population, as this adds much complexity to the model without giving substantial new insights. Most of the existing literature also models large populations as continua (see, e.g., Rubinstein and Wolinsky, 1985; Weibull, 1995; Dixit, 2003; Herold and Kuzmics, 2009; Sakovics and Steiner, 2012; Alger and Weibull, 2013). Kandori (1992) and Ellison (1994) show that large finite populations differ from infinite populations because only the former can induce contagious equilibria. However, as noted by Ellison (1994, p. 578), and as discussed in Section 4.2, these contagious equilibria fail in the presence of a single "crazy" agent who always defects (and likewise in a finite population, although we do not formalize this observation in the paper).

[18] We restrict attention to a fixed number of observed actions to simplify the presentation and the notation. As discussed in Comment 4 in Section 4.3, our results can be extended to a setup in which the number of observed actions is random.

An environment is a pair $E = (G, k)$, where $G = (A, \pi)$ is a two-player symmetric normal-form game, and $k \in \mathbb{N}$ is the number of observed actions. Let $A = \{a_1, ..., a_{|A|}\}$ be the finite set of actions, and let $\pi : A \times A \to \mathbb{R}$ be the payoff function of the underlying game. Let $\Delta(A)$ denote the set of mixed actions (resp., distributions over A), and let π be extended to mixed actions in the usual linear way. We use the letter a (resp., α) to denote a typical pure (mixed) action. With a slight abuse of notation let $a \in A$ also denote the element in $\Delta(A)$ that assigns probability 1 to a. *We adopt this convention for all probability distributions throughout the paper.*

Remark 2. The assumption that the underlying game is symmetric is essentially without loss of generality (if G is played within a single population). Asymmetric games can be symmetrized by considering an extended game in which agents are randomly assigned to the different player positions with equal probability, and the agent's strategy conditions his played action on the assigned role (see, e.g., Selten, 1980).

2.2 Stationary Strategy

The signal observed about the partner is the number of times he played each action $a \in A$ in the sample of k observed actions. Fix an environment $E = ((A, \pi), k)$. Let M denote the set of feasible signals:

$$M = \left\{ m \in \mathbb{N}^{|A|} \middle| \sum_i m_i = k \right\},$$

where m_i is interpreted as the number of times that action a_i is observed in the sample. Given a distribution of actions $\alpha \in \Delta(A)$ and an environment $E = (G, k)$, let $\nu_\alpha(m_1, ..., m_{|A|})$ be the probability of and agent observing signal $(m_1, ..., m_{|A|})$ conditional on being matched with a partner who plays on average the distribution of actions $\alpha \in \Delta(A)$. That is, $\nu(\alpha) := \nu_\alpha \in \Delta(M)$ is a multinomial signal distribution that describes a sample of k i.i.d. actions, where each action is distributed according to α:

$$\forall (m_1, ..., m_{|A|}) \in M, \quad \nu_\alpha(m_1, ..., m_{|A|}) = \frac{k!}{m_1! \cdots m_{|A|}!} \cdot (\alpha(a_1))^{m_1} \cdots \left(\alpha\left(a_{|A|}\right)\right)^{m_{|A|}}. \tag{1}$$

Let $\Delta^{mn}(M)$ denote the set of multinomial signal distributions. That is, a signal distribution ν^* is an element of $\Delta^{mn}(M)$ iff there exists a distribution of actions $\alpha^* \in \Delta(A)$ such that $\nu^* = \nu(\alpha)$. Given $\nu^* \in \Delta^{nm}(M)$, let $\alpha(\nu^*) = \alpha_{\nu^*} \in \Delta(A)$ be the distribution of actions that induce signals distributed according to ν^*, i.e., $\nu(\alpha(\nu^*)) = \nu^*$.

A *stationary strategy* (henceforth, *strategy)* is a mapping $s : M \to \triangle(A)$ that assigns a mixed action to each possible signal. Let $s_m \in \triangle(A)$ denote the mixed action assigned by strategy s after observing signal m. That is, for each action $a \in A$, $s_m(a) = s(m)(a)$ is the probability that a player who follows strategy s plays action a after observing signal m. We also let a denote the strategy s that plays action a regardless of the signal, i.e., $s_m(a) = 1$ for all $m \in M$. Strategy s is *totally mixed*, if for each action $a \in A$, and signal $m \in M$ $s_m(a) > 0$. Let $\mathcal{S}$ denote the set of all strategies. Given strategy s and distribution of signals $\nu \in \Delta(M)$, let $s(\nu) \in \Delta(A)$ be the distribution of actions played by an agent who follows strategy s and observes a signal sampled from ν:

$$\forall a \in A, \quad s(\nu)(a) = \sum_{m \in M} \nu(m) \cdot s_m(a).$$

2.3 Signal Profile and Steady State

Fix environment $(G = (A, \pi), k)$ and finite set of strategies S. A *signal profile* $\theta : S \to \triangle^{mn}(M)$ *is* a function that assigns a multinomial distribution of signals for each strategy in S. Let O_S be the set of all signal profiles

defined over S.

Given a strategy $\sigma \in \Delta(S)$ and a signal profile $\theta \in O_S$, let $\theta_\sigma \in \Delta(M)$ be the *average distribution of signals in the population*, i.e., $\theta_\sigma(m) := \sum_{s \in S} \sigma(s) \cdot \theta_s(m)$.

Let $f_\sigma : O_S \to O_S$ be the *mapping between signal profiles* that is induced by σ. That is, $f_\sigma(\theta)$ is the "new" signal profile that is induced by players who follow strategy distribution σ, and who observe signals about the partners according to the "current" signal profile θ. Specifically, when Alice, who follows strategy s, is being matched with a random partner whose strategy is sampled according to σ, she observes a random signal according to the "current" average distribution of signals in the population θ_σ. As a result her distribution of actions is $s(\theta_\sigma)$, and thus her behavior induces the signal distribution $\nu(s(\theta_\sigma))$. Thus, we define this latter expression as her "new" distribution of signals $(f_\sigma(\theta))_s$. Formally:

$$\forall m \in M,\ s \in S,\quad (f_\sigma(\theta))_s(m) = \nu(s(\theta_\sigma))(m). \tag{2}$$

We say that a signal profile $\theta : S \to \triangle^{mn}(M)$ is c*onsistent* with distribution of strategies σ if it is a fixed point of the mapping $f_\sigma(\theta)$, i.e., if $f_\sigma(\theta) = \theta$. The interpretation of the consistency requirement is that a population of agents who follow the distribution of strategies σ and observe signals about the partners sampled from the profile θ will behave in a way that induces the same profile of signal distributions θ.

A steady state of an environment (G, k) is a triple consisting of (1) a finite set of strategies S interpreted as the strategies that are played by the agents in the population, (2) a distribution σ over S interpreted as a description of the fraction of agents following each strategy, and (3) a consistent signal profile $\theta : S \to \Delta^{mn}(M)$. Formally:

Definition 1. A *steady state* (or *state* for short) of an environment (G, k) is a triple (S, σ, θ) where $S \subseteq \mathcal{S}$ is a finite set of strategies, $\sigma \in \Delta(S)$ is a distribution with full support over S, and $\theta : S \to \triangle^{mn}(M)$ is a consistent signal profile (i.e., $f_\sigma(\theta) = \theta$).

When the set of strategies is a singleton, i.e., $S = \{s\}$, we omit the degenerate distribution assigning a mass of one to s, and we write the steady state as a pair $(\{s\}, \theta)$. *We adopt this convention, of omitting reference to degenerate distributions, throughout the paper.*

A standard fixed-point argument shows that any distribution of strategies admits a consistent signal profile.

Lemma 1. *Let S be a finite set of strategies and let $\sigma \in \Delta(S)$ be a distribution. Then, there exists a* consistent signal profile $\theta : S \to \triangle^{mn}(M)$ *such that (S, σ, θ) is a steady state.*

Proof. Observe that the space O_S is a convex and compact subset of a Euclidean space, and that the mapping $f_\sigma : O_S \to O_S$ (defined in (2) above) is continuous. Brouwer's fixed-point theorem implies that the mapping σ has a fixed point, which is a consistent outcome by definition. □

Some distributions induce multiple consistent profiles of signal distributions. For example, suppose that the underlying game is the Prisoner's Dilemma, each agent observes three of the partner's actions (i.e., $k = 3$), and everyone follows the strategy of playing the most frequently observed action (i.e., using the terminology introduced below, $S = \{s^2\}$, where $s^2(m) = d$ iff $m \geq 2$). In this setting there are three consistent profiles of signal distributions: one in which everyone cooperates (i.e., $\theta_{s^2} = \nu_c$), one in which everyone defects (i.e., $\theta_{s^2} = \nu_d$), and one in which everyone plays (on average) uniformly[19] (i.e., $\theta_{s^2} = \nu_{(0.5 \cdot d + 0.5 \cdot c)}$ where we let $(0.5 \cdot d + 0.5 \cdot c)$ denote the distribution that puts equal probability on each of the two actions).

[19] In Heller and Mohlin (2016b) we study a setup in which the number of observed actions is random, and we show that all strategy distributions admit unique consistent profiles of signal distributions iff the expected number of observed actions is less than one.

Remark 3. In order for a steady state to be a plausible candidate for a long-run outcome it is natural to require the steady state to be dynamically stable in the sense that if the signal profile is perturbed slightly then this does not induce a movement away from the steady state. More precisely, we may say that a steady state $(S^*, \sigma^*, \theta^*)$ is Lyapunov stable if for each $\epsilon > 0$, there exists a $\delta > 0$, such that if one perturbs the initial distribution of signals from θ^* to a δ-nearby profile θ, then the distribution of signals in the population remains within distance ϵ of θ^*, i.e., $f_\sigma^n(\theta)$is ϵ-nearby to θ^* for each n. In Section 3.3 we define a related notion of robustness to signal perturbations (see, Definition 8), and we show that all the equilibria presented in our main results satisfy this refinement.

2.4 Perturbed Environment

In a seminal paper Kreps, Milgrom, Roberts, and Wilson (1982) show, in a standard setup of a two-player finitely repeated Prisoner's Dilemma, that the equilibrium analysis completely changes if one slightly perturbs the environment by assuming that with very small probability one of the players may be committed to following a "tit-for-tat" strategy. (See Mailath and Samuelson, 2006, for a textbook analysis and a survey of the "reputation" literature.) Motivated by this observation, we introduce a notion of perturbed environment in which a small fraction of agents in the population are committed to playing specific strategies, even though these strategies are not necessarily payoff-maximizing.

A perturbed environment is a tuple consisting of (1) an environment, (2) a distribution λ over a set of commitment strategies S^C that includes a totally mixed strategy, and (3) a number ϵ representing the share of agents who are committed to playing strategies in S^C (henceforth, *committed agents*). The remaining $1-\epsilon$ share of the agents can play any strategy in $\mathcal{S}$ (henceforth, *normal agents*). Formally:

Definition 2. A *perturbed environment* is a tuple $E_\epsilon = \left((G,k), \left(S^C, \lambda\right), \epsilon\right)$, where G is the underlying game, $k \in \mathbb{N}$ is the number of observed actions, S^C is a non-empty finite set of strategies (called, *commitment strategies*) that includes a totally mixed strategy, $\lambda \in \Delta\left(S^C\right)$ is a distribution with full support over the commitment strategies, and $\epsilon \geq 0$ is the mass of committed agents in the population.

We require S^C to include at least one totally mixed strategy because we want all signals to be observed with positive probability in a perturbed environment when $\epsilon > 0$. (This is analogous to the requirement in Selten, 1975, that all actions be played with positive probability in the perturbations defining a perfect equilibrium.)

We refer to $\left(S^C, \lambda\right)$ as a *distribution of commitments*. With a slight abuse of notation, we identify an *unperturbed environment* $\left((G,k), \left(S^C, \lambda\right), \epsilon = 0\right)$ with the equivalent environment (G,k).

Remark 4. To simplify the presentation, the definition of perturbed environment includes only commitment strategies, and it does not allow "trembling hand" mistakes. As discussed in Section 7.3, the results also hold in a setup in which agents also tremble, as long as the probability by which a normal agent trembles is of the same order of magnitude as the frequency of committed agents.

One of our main results (Theorem 1) requires an additional mild assumption on the perturbed environment that rules out the knife-edge case in which all agents (committed and non-committed alike) behave exactly the same. Specifically, a set of commitments is regular if for each distribution of actions α, there exists a committed strategy s that does not play distribution α when observing the signal distribution induced by α. Formally:

Definition 3. A set of commitment strategies S^C is *regular* if for each distribution of actions $\alpha \in \Delta(A)$, there exists a strategy $s \in S^C$ such that $s_{\nu(\alpha)} \neq \alpha$.

If the set of commitments is regular, then we say that the distribution $\left(S^C, \lambda\right)$ and the perturbed environment $\left((G,k), \left(S^C, \lambda\right), \epsilon\right)$ are regular. An example of a regular set of commitments is the set that includes two strategies $s \equiv \alpha_1$ and $s' \equiv \alpha_2$ that induce agents to play mixed actions $\alpha_1 \neq \alpha_2$ regardless of the observed signal.

2.5 Steady State in a Perturbed Environment

We now adapt the definitions of a consistent signal profile and of a steady state to perturbed environments.

Fix a perturbed environment $E_\epsilon = \left((G,k), \left(S^C, \lambda\right), \epsilon\right)$ and a finite set of strategies S^N, interpreted as the strategies followed by the normal agents in the population. We redefine a *signal profile* $\theta : S^C \cup S^N \to \triangle^{mn}(M)$ as a function that assigns a multinomial distribution of signals to each strategy in $S^C \cup S^N$. Let $O_{S^C \cup S^N}$ be the set of all signal profiles defined over $S^C \cup S^N$.

Given a distribution over strategies of the normal agents $\sigma \in \Delta\left(S^N\right)$ and a signal profile $\theta \in O_{S^C \cup S^N}$, let $\theta_{((1-\epsilon)\cdot\sigma+\epsilon\cdot\lambda)} \in \Delta^{mn}(M)$ be the *average distribution of signals in the population*, i.e., $\theta_{((1-\epsilon)\cdot\sigma+\epsilon\cdot\lambda)}(m) := \sum_{s \in S^C \cup S^N} \left((1-\epsilon)\cdot\sigma + \epsilon\cdot\lambda\right)(s) \cdot \theta_s(m)$, and let $\theta_\sigma \in \Delta^{mn}(M)$ be the *average distribution of signals among the normal agents*, i.e., $\theta_\sigma(m) := \sum_{s \in S^N} \sigma(s) \cdot \theta_s(m)$.

Let $f_{((1-\epsilon)\cdot\sigma+\epsilon\cdot\lambda)} : O_S \to O_S$ be the *mapping between signal profiles* that is induced by the population's distribution over strategies $((1-\epsilon)\cdot\sigma + \epsilon\cdot\lambda)$. That is, $f_{((1-\epsilon)\cdot\sigma+\epsilon\cdot\lambda)}(\theta)$ is the "new" signal profile that is induced by a population of normal agents who follow strategy distribution σ and committed agents who follow strategy distribution λ, and who observe signals about the partners according to the "current" signal profile θ. Specifically, when Alice, who follows strategy s, is being matched with a random partner whose strategy is sampled according to $(1-\epsilon)\cdot\sigma + \epsilon\cdot\lambda$, she observes a random signal according to the "current" average distribution of signals in the population $\theta_{((1-\epsilon)\cdot\sigma+\epsilon\cdot\lambda)}$. As a result her distribution of actions is $s\left(\theta_{((1-\epsilon)\cdot\sigma+\epsilon\cdot\lambda)}\right)$, and consequently her behavior induces the signal distribution $\nu\left(s\left(\theta_{((1-\epsilon)\cdot\sigma+\epsilon\cdot\lambda)}\right)\right)$. Thus, we define this latter expression as her "new" distribution of signals $\left(f_{((1-\epsilon)\cdot\sigma+\epsilon\cdot\lambda)}(\theta)\right)_s$. Formally:

$$\forall m \in M,\ s \in S,\quad \left(f_{((1-\epsilon)\cdot\sigma+\epsilon\cdot\lambda)}(\theta)\right)_s(m) = \nu\left(s\left(\theta_{((1-\epsilon)\cdot\sigma+\epsilon\cdot\lambda)}\right)\right)(m). \tag{3}$$

Given a distribution of strategies $((1-\epsilon)\cdot\sigma + \epsilon\cdot\lambda)$, we say that a signal profile $\theta^* : S^C \cup S^N \to \triangle^{mn}(M)$ is *consistent* if it is a fixed point of the mapping $f_{((1-\epsilon)\cdot\sigma+\epsilon\cdot\lambda)}$, i.e., if $f_{((1-\epsilon)\cdot\sigma+\epsilon\cdot\lambda)}(\theta^*) = \theta^*$.

Finally, we adapt the definition of a steady state as follows:

Definition 4. A *steady state* (or *state* for short) of a perturbed environment $\left((G,k), \left(S^C, \lambda\right), \epsilon\right)$ is a triple $\left(S^N, \sigma, \theta\right)$ where $S^N \subseteq \mathcal{S}$ is a finite set of strategies (called, *normal strategies*) , $\sigma \in \Delta\left(S^N\right)$ is a distribution with a full support over S^N, and $\theta : S^N \cup S^C \to \triangle^{mn}(M)$ is a consistent signal profile.

The following example demonstrates a specific steady state in a specific perturbed environment in the Prisoner's Dilemma game. The example is intended to clarify the various definitions of this section and, in particular, the consistency requirement. Later, we revisit the same example to explain the essentially unique perfect equilibrium that supports cooperation.

Example 1. Consider the perturbed environment $\left((G_{PD}, 2), \left(\{s^u \equiv 0.5\}\right), \epsilon\right)$, in which the underlying game is the Prisoner's Dilemma, each agent observes two of her partner's actions, there is a single commitment strategy, denoted by s^u, which is followed by a fraction $0 < \epsilon << 1$ of committed agents, who choose each action with probability 0.5 regardless of the observed signal. Let $\left(S = \left\{s^1, s^2\right\}, \sigma = \left(\frac{1}{6}, \frac{5}{6}\right), \theta\right)$ be the following steady state. The state includes two normal strategies: s^1 and s^2. The strategy s^1 defects iff $m \geq 1$, and the strategy s^2

defects iff $m \geq 2$. The distribution σ assigns a mass of $\frac{1}{6}$ to s^1 and a mass of $\frac{5}{6}$ to s^2. The consistent signal profile θ is defined as follows (neglecting terms of $O\left(\epsilon^2\right)$ throughout the example):

$$\theta_{s^u}(m) = \begin{cases} 25\% & if\ m = 2 \cdot c \\ 50\% & if\ m = c, d \\ 25\% & if\ m = 2 \cdot d, \end{cases} \quad \theta_{s^1}(m) = \begin{cases} 1 - 3.5 \cdot \epsilon & if\ m = 2 \cdot c \\ 3.5 \cdot \epsilon & if\ m = c, d \\ O\left(\epsilon^2\right) & if\ m = 2 \cdot d \end{cases} \quad \theta_{s^2}(m) = \begin{cases} 1 - 0.5 \cdot \epsilon & if\ m = 2 \cdot c \\ 0.5 \cdot \epsilon & if\ m = c, d \\ O\left(\epsilon^2\right) & if\ m = 2 \cdot d. \end{cases} \tag{4}$$

To confirm the consistency of θ, we have first to calculate the average distribution of signals in the population:

$$\theta_{((1-\epsilon)\cdot\sigma+\epsilon\cdot\lambda)}(m) = \begin{cases} 1 - 1.75 \cdot \epsilon & if\ m = 2 \cdot c \\ 1.5 \cdot \epsilon & if\ m = c, d \\ 0.25 \cdot \epsilon & if\ m = 2 \cdot d. \end{cases}$$

Using $\theta_{((1-\epsilon)\cdot\sigma+\epsilon\cdot\lambda)}$, we confirm the consistency of θ_{s^1} and θ_{s^2} by showing that $\theta_{s^i} = \nu\left(s^i\left(\theta_\sigma\right)\right)$ (the consistency of $\theta_{s^u}$is immediate). We do so by calculating distribution of actions played by a player following strategy s_i who observes the distribution of actions of a random partner:

$$\begin{array}{rclrcl} s^1\left(\theta_{((1-\epsilon)\cdot\sigma+\epsilon\cdot\lambda)}\right)(c) & = & 1 - 1.75 \cdot \epsilon & s^2\left(\theta_{((1-\epsilon)\cdot\sigma+\epsilon\cdot\lambda)}\right)(c) & = & 1 - 0.25 \cdot \epsilon, \\ s^1\left(\theta_{((1-\epsilon)\cdot\sigma+\epsilon\cdot\lambda)}\right)(d) & = & 1.75 \cdot \epsilon & s^2\left(\theta_{((1-\epsilon)\cdot\sigma+\epsilon\cdot\lambda)}\right)(d) & = & 0.25 \cdot \epsilon. \end{array}$$

Note that $s^1\left(\theta_{((1-\epsilon)\cdot\sigma+\epsilon\cdot\lambda)}\right)(d) = 1 - \theta_{((1-\epsilon)\cdot\sigma+\epsilon\cdot\lambda)}(2 \cdot c)$ and $s^2\left(\theta_{((1-\epsilon)\cdot\sigma+\epsilon\cdot\lambda)}\right)(d) = \theta_{((1-\epsilon)\cdot\sigma+\epsilon\cdot\lambda)}(2 \cdot d)$. The final step in showing that θ is a consistent profile is the observation that each θ_{s^i} coincides with the multinomial distribution that is induced by $s^i\left(\theta_{((1-\epsilon)\cdot\sigma+\epsilon\cdot\lambda)}\right)$.

Remark 5. The argument in Example 1 ignores the small terms of $O\left(\epsilon^2\right)$. The steady state in the above example is the essentially unique state that induces full cooperation in the main results of the paper, and we later present a general argument that shows that indeed the small terms of $O\left(\epsilon^2\right)$ do not interfere with the steady state (and that moreover, starting from any slightly perturbed initial distribution of signals, they will converge to this steady states).

2.6 Discussion of the Model

Our model differs from most of the existing literature on community enforcement in three key dimensions (see, e.g., Kandori, 1992; Ellison, 1994; Dixit, 2003; Takahashi, 2010; Deb, 2012; Deb and González-Díaz, 2014). In what follows we discuss these three key differences, and their implications on our results.

1. *The presence of a few committed agents.* The existing reputation literature includes various models in which a patient long-run agent plays a repeated game, and there is a small probability of the agent (she) being a commitment type (e.g., Kreps, Milgrom, Roberts, and Wilson, 1982; Fudenberg and Levine, 1989; Celetani, Fudenberg, Levine, and Pesendorfer, 1996; see Mailath and Samuelson, 2006, for a textbook analysis and survey). In most of this literature the agent's partners can observe all of her past behavior, and this yields the main stylized result obtained in these reputation models, namely, that there exists a high *lower* bound on all (Nash) equilibrium payoffs.[20] To the best of our knowledge, we are the first to

[20] One notable exception is Ely, Fudenberg, and Levine (2008), which studies games between a long-run player and a sequence of short-run players. They show that if the participation of the short-run players is optional, and if every action of the long-run player that makes the short-run players want to participate can be interpreted as a signal that the long-run player is "bad," then

introduce commitment types to the community enforcement setup in which agents are randomly matched, and each agent can observe only a few past actions of the partner. The role played by the commitment types in our setup is substantially different from their role in the existing reputation literature.[21] If one removes the commitment types from our setup, then one can show (by using belief-free equilibria, as in Takahashi, 2010) that: (1) it is always possible to support full cooperation as an equilibrium outcome, and (2) there are various strategies that sustain full cooperation.

The results of this paper show that the introduction of a few committed agents, regardless of how they behave, implies very different results: (1) a *very low* equilibrium payoff is obtained in offensive Prisoner's Dilemmas (Theorem 1), and (2) there is an essentially unique strategy combination that supports a cooperative equilibrium in defensive Prisoner's Dilemmas. The reason for these differences is that the presence of committed agents implies that the population is heterogeneous regardless of the equilibrium behavior, and hence the observation of past actions must have some influence on the likely behavior of the partner in the current match (more detailed discussions of this issue follow Theorem 1 and Remark 10).

2. *Restriction to Stationary Strategies.* In our model we restrict agents to using stationary strategies that condition only on the number of times they observed each of the partner's actions being played in past interactions (or on the number of observed action profiles, as in the extended model of Section 5). We do not allow agents to condition their play on the order in which the observed actions were played in the past, nor on the agent's own history of play, nor on calendar time. The assumption is made for tractability. It allows us to simplify the presentation of the model and results. In addition, the assumption strengthens our results by allowing us to achieve two kinds of uniqueness results that do not hold without stationarity. First, it yields the result that always defecting is the unique perfect equilibrium in an offensive Prisoner's Dilemma (Theorem 1); without this assumption we can show only the weaker result that full cooperation is not a perfect equilibrium outcome in offensive Prisoner's Dilemmas (Theorem 6). Second, it yields the result that there is essentially a unique strategy that supports full cooperation, whenever full cooperation can be supported; without this assumption, there might be additional strategies that support full cooperation. Additionally, the stationarity assumption allows the cooperative outcome to be *strictly* perfect, i.e., to be the limit of Nash equilibrium outcomes in any converging sequence of perturbed environments.[22]

3. *Not having a "global time zero."* Most of the existing literature represents interactions within a community as a repeated game that has a "global time zero," in which the first ever interaction takes place. In many real-life situations, the interactions within a community began a long time ago and have continued, via overlapping generations, to the present day. It seems implausible that today's agents condition their behavior on what happened in the remote past (or on calendar time). For example, trade interactions have been been taking place from time immemorial. It seems unreasonable to assume that Alice's behavior

reputation is "bad" in the sense that it uniquely chooses a low equilibrium payoff to the long-run player. Another exception is Cripps, Mailath, and Samuelson (2004), which shows that a normal agent cannot indefinitely sustain a reputation for non-credible committed behavior.

[21] In addition, there is a difference in the "order of limits." The existing literature typically looks at the set of Nash equilibrium payoffs, given a fixed small frequency of committed agents, when the discount factor converges to one. Contrary to this approach, we look at a fixed discount factor (equal to one in the baseline model, and to a fixed high value less than one in the model of Section 6), when the frequency of committed agents converges to zero, as is typically done in the refinement literature (e.g., in the notion of "perfect equilibrium" of Selten, 1975).

[22] Bhaskar, Mailath, and Morris (2013) presents a theoretical foundation for focusing on stationary equilibria, albeit in a substantially different setup. Specifically, they study repeated games in which agents interact sequentially and have bounded memory, and they show that any equilibrium that satisfies a refinement à la Harsanyi's purification must be stationary.

today is conditioned on what transpired in some long-forgotten time $t = 0$, when, say, two hunter-gatherers were involved in the first ever trade. We suggest that, even though real-world interactions obviously begin at some definite date, the best way of modeling what the interacting agents think about the situation may be to get rid of global time zero and focus on strategies that do not condition on what happened in the remote past, or on calendar time.[23] The lack of a global time zero is the reason why, unlike in repeated games, a distribution of strategies does not uniquely determine the behavior and the payoffs of the agent, so that one must explicitly add the consistent signal profile θ as part of the description of the state of the population.

It is possible to interpret a steady state (S, σ, θ) as a kind of initial condition for society, in which agents already have a long-existing past. That is, we begin our analysis of community interaction at a point in time when agents have for a long time followed the strategy distribution (S, σ) yielding the consistent signal profile θ. We then ask whether any agent has a profitable deviation from her strategy. If not, then the steady state (S, σ, θ) is likely to persist (in Appendix B we discuss robustness to small perturbations and to joint deviations by a small group of agents). This approach stands in contrast to the standard approach that studies whether or not agents have a profitable deviation at a time $t >> 1$ following a long history that started with the first ever interaction at $t = 0$.

In Section 6 we present a conventional repeated game model that differs from the existing literature in only one key aspect: the presence of a few committed agents. In particular, this alternative model features standard calendar time (i.e., the community starts interacting at time zero), and agents discount the future, and are not limited to choosing only stationary strategies. We show that most of our results, or qualitatively similar results, hold also in this setup (except the uniqueness results discussed above, which are no longer valid due to our omitting the limitation to stationary strategies). We feel that this alternative model, while being closer to the existing literature than the main model, suffers from added technical complexity that may hinder the model from being insightful and accessible.

3 Solution Concept

3.1 Long-Run Payoff

In this subsection we define the long-run average (per-round) payoff of a patient agent who follows a stationary strategy s, given a steady state $\left(S^N, \sigma, \theta\right)$ of a perturbed environment $\left((G,k), \left(S^C, \lambda\right), \epsilon\right)$. The same definition, when taking $\epsilon = 0$, holds for an unperturbed environment.

We begin by extending the definition of a consistent signal profile θ to non-incumbent strategies. For each non-incumbent strategy $\hat{s} \in \mathcal{S} \backslash \left(S^N \cup S^C\right)$, define $\theta\left(\hat{s}\right) = \theta_{\hat{s}}$ as the distribution of signals induced by a deviating agent who follows strategy $\hat{s}$ and observes the distribution of signals induced by a random partner in the population (sampled according to $(1-\epsilon)\cdot\sigma\left(s'\right) + \epsilon \cdot \lambda\left(s'\right)$). That is, for each strategy $\hat{s} \in \mathcal{S} \backslash \left(S \cup S^C\right)$, and each signal $m \in M$, we define

$$\theta_{\hat{s}}\left(m\right) = \left(\nu\left(\hat{s}\left(\theta_{\left((1-\epsilon)\cdot\sigma + \epsilon\cdot\lambda\right)}\right)\right)\right)\left(m\right).$$

[23] For a related discussion of the proper way of modeling finitely and infinitely repeated interactions, see Osborne and Rubinstein (1994, p. 135).

We define the long-run payoff of an agent who follows an arbitrary strategy $s \in \mathcal{S}$ as:

$$\pi_s\left(S^N, \sigma, \theta\right) = \sum_{s' \in S^N \cup S^C} \left((1-\epsilon)\cdot\sigma\left(s'\right) + \epsilon\cdot\lambda\left(s'\right)\right)\cdot\left(\sum_{(a,a')\in A\times A} s_{\theta(s')}\left(a\right)\cdot s'_{\theta(s)}\left(a'\right)\cdot\pi\left(a,a'\right)\right). \tag{5}$$

Eq. (5) is straightforward. The inner (right-hand) sum (i.e., $\sum_{(a,a')\in A\times A} s_{\theta(s')}\left(a\right)\cdot s'_{\theta(s)}\left(a'\right)\cdot\pi\left(a,a'\right)$) calculates the expected payoff of Alice who follows strategy s conditional on being matched with a partner who follows strategy s'. The outer sum weighs these conditional expected payoffs according to the frequency of each incumbent strategy s' (i.e., $\left((1-\epsilon)\cdot\sigma\left(s'\right)+\epsilon\cdot\lambda\left(s'\right)\right)$), which yields the expected payoff of Alice against a random partner in the population.

Let $\pi\left(S,\sigma,\theta\right)$ be the average payoff of the *normal* agents in the population:

$$\pi\left(S^N,\sigma,\theta\right) = \sum_{s\in S^N}\sigma\left(s\right)\cdot\pi_s\left(S^N,\sigma,\theta\right).$$

3.2 Nash and Perfect Equilibrium

A steady state is a Nash equilibrium if no agent can obtain a higher payoff by a unilateral deviation. Formally:

Definition 5. The steady state $\left(S^N,\sigma,\theta\right)$ of perturbed environment $\left(\left(G,k\right),\left(S^C,\lambda\right),\epsilon\right)$ is a *Nash equilibrium* if for each strategy $s\in\mathcal{S}$, it is the case that $\pi_s\left(S^N,\sigma,\theta\right)\leq\pi\left(S^N,\sigma,\theta\right)$.

Note that the $1-\epsilon$ normal agents in such a Nash equilibrium must obtain the same maximal payoff. That is, each normal strategy $s\in S^N$ satisfies $\pi_s\left(S^N,\sigma,\theta\right)=\pi\left(S^N,\sigma,\theta\right)\geq\pi_{s'}\left(S^N,\sigma,\theta\right)$ for each strategy $s'\in\mathcal{S}$. However, the ϵ committed agents may obtain lower payoffs.

Next, observe that any symmetric Nash equilibrium (α,α) of the underlying game can be implemented in a corresponding Nash equilibrium of the unperturbed environment in which everyone plays α regardless of the observed signal.

Fact 1. *Let $\alpha\in\Delta\left(A\right)$ be a symmetric Nash equilibrium strategy of the underlying game $G=\left(A,\pi\right)$. Then the steady state $\left(S^N=\left\{\alpha\right\},\alpha\right)$ in which everyone plays α regardless of the observed signal is a Nash equilibrium in the unperturbed environment (G,k) for any $k\in\mathbb{N}$.*

A steady state is a (regular) perfect equilibrium if it is the limit of Nash equilibria of (regular) perturbed environments when the frequency of the committed agents converges to zero. Formally, starting with standard definitions of convergence of a sequence of strategies and of a sequence of states, we have:

Definition 6 (Convergence of Strategies, Distributions, and States)**.** Fix environment (G,k). A sequence of strategies $\left(s_n\right)_n$ converges to strategy s (denoted by $\left(s_n\right)_n\to_{n\to\infty}s$) if for each signal $m\in M$ and each action a, the sequence of probabilities $\left(s_n\right)_m\left(a\right)$ converges to $s_m\left(a\right)$. A distribution of signals $\left(\nu_n\right)_n$ converges to ν (denoted by $\left(\nu_n\right)_n\to_{n\to\infty}\nu$) if the sequence of probabilities $\left(\nu_n\right)\left(m\right)$ converges to $\nu\left(m\right)$ for each signal m. A sequence of states $\left(S_n^N,\sigma_n,\theta_n\right)_n$ converges to a state $\left(S^*,\sigma^*,\theta^*\right)$ if for each strategy $s\in supp\left(\sigma^*\right)$, there exists a sequence of sets of strategies $\left(\hat{S}_n^N\right)_n$, with $\hat{S}_n^N\subseteq S_n^N$ for each n, such that (1) $\sum_{s_n\in\hat{S}_n^N}\sigma_n\left(s_n\right)\to\sigma^*\left(s\right)$, and for each sequence of elements of those sets (i.e., for each sequence of strategies $\left(s_n\right)_n$ such that $s_n\in\hat{S}_n^N$ for each n), (2) $s_n\to_{n\to\infty}s$, and (3) $\theta_n\left(s_n\right)\to\theta^*\left(s\right)$.

Definition 7. A steady state $\left(S^*,\sigma^*,\theta^*\right)$ of the environment (G,k) is a *(regular) perfect equilibrium* if there exist a (regular) distribution of commitments $\left(S^C,\lambda\right)$ and converging sequences $\left(S_n^N,\sigma_n,\theta_n\right)_n\to_{n\to\infty}\left(S^*,\sigma^*,\theta^*\right)$

and $(\epsilon_n > 0)_n \to_{n\to\infty} 0$, such that for each n, the state $\left(S_n^N, \sigma_n, \theta_n\right)$ is a Nash equilibrium of the perturbed environment $\left((G,k), \left(S^C, \lambda\right), \epsilon_n\right)$. In this case, we say that $(S^*, \sigma^*, \theta^*)$ is a *perfect equilibrium with respect to distribution of commitments* $\left(S^C, \lambda\right)$. If $\theta^* \equiv a$, we say that action $a \in A$ is a *perfect* equilibrium action.

By standard continuity arguments, any perfect equilibrium is a Nash equilibrium of the unperturbed environment. Next, observe that any symmetric "trembling-hand" perfect equilibrium (Selten, 1975) α of the underlying game corresponds to a perfect equilibrium of the environment in which all normal agents play α regardless of the observed signal. Moreover, if α is not totally mixed, then this perfect equilibrium is a regular perfect equilibrium. Formally:

Proposition 1. *Let $\alpha \in A$ be a symmetric perfect equilibrium action of the underlying game $G = (A, \pi)$. Then the state $(S = \{\alpha\}, \nu_\alpha)$ is a perfect equilibrium in the environment (G, k) for any $k \in \mathbb{N}$. Moreover, if the distribution α is not totally mixed, then $\left(S^N = \{\alpha\}, \nu_\alpha\right)$ is a regular perfect equilibrium.*

An underlying game $G = ((a,b), \pi)$ is a *(two-action) coordination game* if (a,a) and (b,b) are strict Nash equilibria. The next result shows that the totally mixed equilibrium of such a game does not correspond to a regular perfect equilibrium in any environment with[24] $k \geq 1$.

Proposition 2. *Let $G = (\{a,b\}, \pi)$ be a coordination game. Let $\alpha \in \Delta(\{a,b\})$ be the mixed equilibrium action of G. Then the state $\left(S^N = \{\alpha\}, \nu_\alpha\right)$ is not a regular perfect equilibrium in (G,k) for any $k \geq 1$.*

The intuition is that in the mixed equilibrium both actions earn the same expected payoff. The regularity of the set of commitment strategies implies there exists action a such that when an agent observes a sequence of only a:s, then the unique best reply is to play a because the partner is more likely to play a as well.

3.3 Robust Strictly Perfect Equilibrium Action

In this section we define two refinements of the notion of a perfect equilibrium: strictness and robustness.

In most of our results we focus on pure perfect equilibria, in which there exists action a^* that is played with probability one in the limit in which the frequency of committed agents converges to zero. In order to simplify the notation, we define the following refinements only with respect to pure equilibrium outcomes. The notion of perfect equilibrium may be considered too weak due to two issues:

1. A perfect equilibrium may crucially depend on a specific distribution of commitment strategies. We solve this issue by adapting to the current setup the refinement of strict perfection (Okada, 1981), which requires the pure equilibrium outcome to be the limit behavior of Nash equilibria with respect to *all* commitment strategies.[25]

2. The equilibrium outcome may be unstable in the sense that small perturbations of the distribution of observed signals may induce a change of behavior that moves the population away from the consistent signal profile. We address this issue by introducing a robustness refinement (in the spirit of the notion of Lyapunov stability in dynamic environments) that requires that if we slightly perturb the distribution of observed signals, then the agents still play the pure equilibrium outcome a^* with a probability very close

[24] The result can be extended to the totally mixed equilibrium of a coordination game with more than two actions.

[25] Okada (1981) deals with normal-form games and presents the related notion of a strict perfect equilibrium as the limit of Nash equilibria for any "trembling-hand" perturbation. In our setup different strategies might be equivalent in the sense that they induce the same observable behavior, as the frequency of the commitment agents converges to zero. Our notion focuses on the observed behavior (i.e., everyone playing action a^*), but allows for the choice of strategy that induces the pure action a^* to depend on the distribution of commitments. This approach is in the spirit of other set-wise solution concepts in the literature, such as evolutionarily stable sets (Thomas, 1985) and hyperstable sets (Kohlberg and Mertens, 1986).

to one. Specifically, we require that for any distribution of commitment strategies there exist a bounded sequence of parameters κ_n such that for each perturbed environment with ϵ_n committed agents: (1) the normal agents play action $a^*$with a probability larger than $1-\kappa_n\cdot\epsilon_n$ in the steady state, and (2) if one perturbs the initial distribution of signals to any other (possibly inconsistent) signal profile in which the normal agents are observed to play action $a^*$with a probability of at least $1-\kappa_n\cdot\epsilon_n$, then, agents continue to play action $a^*$with a probability of at least $1-\kappa_n\cdot\epsilon_n$ in the new signal profile that is induced by the agents' behavior and the perturbed signal profile.

Recall that $\alpha(\theta_s)\in\Delta(A)$ is the distribution of actions that induce signals distributed according to $\theta_s\in\Delta^{nm}(M)$, i.e., $\nu(\alpha(\theta_s))=\theta_s$ (as defined in Section 2.2 above). Given a distribution of normal strategies $\left(S^N,\sigma\right)$ and a (possibly inconsistent) signal profile θ, let $\alpha_\sigma(\theta)\in\Delta(A)$ be the (σ-weighted) population average of the distributions of actions that induce signals distributed according to the signal profile $\theta\in\Delta^{nm}(M)$ for the normal agents, i.e., for each action $a\in A$,

$$\alpha_\sigma(\theta)(a)=\sum_{s\in S^N}\sigma(s)\cdot\alpha(\theta_s)(a).$$

That is, $(\alpha(\theta_s))_{s\in S^N}$ is the profile of distributions of actions that generate the profile of signal distributions for the normal agents $\theta=\{\theta_s\}_{s\in S^N}$, and $\alpha_\sigma(\theta)$ is the (σ-weighted) average of the distributions of actions in this profile.

The formal definition of robust strict perfection is as follows.

Definition 8. Action $a^*\in A$ is a *strictly perfect* equilibrium action in the environment $E=(G,k)$ if, for any distribution of commitment strategies $\left(S^C,\lambda\right)$, there exist a steady state $(S^*,\sigma^*,\theta^*\equiv\nu_{a^*})$ and converging sequences $\left(S_n^N,\sigma_n,\theta_n\right)_n\to_{n\to\infty}(S^*,\sigma^*,\theta^*)$ and $(\epsilon_n>0)_n\to_{n\to\infty}0$, such that for each n, the state $\left(S_n^N,\sigma_n,\theta_n\right)$ is a Nash equilibrium of the perturbed environment $\left((G,k),\left(S^C,\lambda\right),\epsilon_n\right)$.
The strictly perfect equilibrium action $a^*\in A$ is *robust* if, in addition, there exists $\kappa>0$ and a sequence $0<(\kappa_n)_n<\kappa$, such that for each n, (1) $\alpha_{\sigma_n}(\theta_n)(a^*)>1-\kappa_n\cdot\epsilon_n$, and (2) for each signal profile $\theta\in O_{(S_n^N\cup S^C)}$,

$$\alpha_{\sigma_n}(\theta)(a^*)\geq 1-\kappa_n\cdot\epsilon_n\ \Rightarrow\ \alpha_\sigma\left(f_{((1-\epsilon)\cdot\sigma_n+\epsilon\cdot\lambda)}\theta\right)(a^*)>1-\kappa_n\cdot\epsilon_n.$$

The following result shows that in an environment in which $k\geq 2$ and in which the game is a two-action coordination game, there is a unique strictly perfect equilibrium action, namely, the Pareto-efficient strict equilibrium action of the underlying game. This holds even if the Pareto-inefficient equilibrium is risk-dominant.[26,27,28]

Proposition 3. *Let (G,k) be an environment where $G=((a,b),\pi)$ is a coordination game and $k\geq 2$. The action a is a strictly perfect equilibrium action in the environment (G,k) if $\pi(a,a)>\pi(b,b)$, and it is not strictly perfect if*[29] *$\pi(a,a)<\pi(b,b)$.*

The essentially unique steady state that supports the Pareto-efficient action as a strictly perfect equilibrium action is similar to the steady-state supporting cooperation in the defensive Prisoner's Dilemma in Theorem 2. It will be presented and discussed in Section 4.

[26] One can show that when $k=1$ both pure actions are strictly perfect.
[27] The formal result deals with coordination games with two actions, but it can be extended to coordination games with more than two actions.
[28] One can adapt the arguments of the robustness result in the proof of Theorem 2 to show that the Pareto-efficient action is also robust (omitted for brevity).
[29] In order to simplify the proof, we restrict attention to almost all commitment strategies (see Remark 11 in the proof). The proof can be extended to all commitment strategies, but as this extension would make the proof much lengthier, we omit it.

The reason why the Pareto-dominated action (say, action a) is not a strictly perfect equilibrium action is the following. Consider a distribution of commitments that includes a commitment strategy that plays action b with high probability. Suppose all normal agents play action a with high probability. This means that if an agent observes a partner always to have played b then the partner is highly likely to be a commitment type who will continue to play b and hence the best response for a normal agent who receives a signal of all b's is to play b. This implies that a deviator who always plays b induces all normal agents to play b, and thus she achieves a payoff of $\pi(b,b)$, which is strictly higher than the incumbents' average payoff (which is close to $\pi(a,a)$).

4 Analysis of the Prisoner's Dilemma

4.1 The Prisoner's Dilemma

In this section we focus on environments in which the underlying game is the Prisoner's Dilemma (denoted by G_{PD}), which is described in Table 3. The class of Prisoner's Dilemma games is fully described by two positive parameters g and l. When both players play action c (*cooperate*) they both get a high payoff (normalized to one), and when they both play action d (*defect*) they both get a low payoff (normalized to zero). When a single player defects he obtains a payoff of $1+g$ (i.e., an additional payoff of g) while his opponent gets $-l$.

Table 3: Matrix Payoffs of Prisoner's Dilemma Games

	c	d
c	1, 1	$-l$, $1+g$
d	$1+g$, $-l$	0, 0

Prisoner's Dilemma
G_{PD}: $g, l > 0$, $g < l+1$

	c	d
c	1, 1	-3, 2
d	2, -3	0, 0

Ex. 1: Defensive PD
G_D: $1 = g < l = 3$

	c	d
c	1, 1	-1.7, 3.3
d	3.3, -1.7	0, 0

Ex. 2: Offensive PD
G_O: $2.3 = g > l = 1.7$

The fact that the Prisoner's Dilemma has two actions allows us to simplify the notation by setting $M = \{0, ..., k\}$, and interpreting $m \in M$ as the number of times that the partner defected in the sampled k observations.

Following Dixit (2003) we classify Prisoner's Dilemma games into two kinds: offensive and defensive.[30] In an *offensive* Prisoner's Dilemma there is a stronger incentive to defect against a cooperator than against a defector (i.e., $g > l$); in a *defensive* PD the opposite holds (i.e., $l > g$). If cooperating is interpreted as exerting high effort, then the defensive PD exhibits strategic complementarity; increasing one's effort from low to high is less costly if the opponent exerts high effort. To see this, consider the following example.

Example 2. Consider a joint project of cowriting an academic paper in which each author can choose either to work hard (cooperate) or to shirk (defect). Assume that the joint paper is accepted (rejected) by a top journal for sure if both authors work hard (shirk), and that the paper is accepted with probability p if a single author works hard. Assume that a publication in a top journal yields a benefit of 6 to each author, and working hard costs 5. If $p < 50\%$ the induced game is defensive (see the game G_D in Table 3 for the case where $p = \frac{1}{3}$), while $p > 50\%$ induces an offensive game (see the game G_O in Table 3 for the case where $p = 55\%$).

[30] Takahashi (2010) calls offensive (defensive) Prisoner's Dilemmas submodular (supermodular).

4.2 Stability of Defection

We begin by showing that defection is strictly perfect in any Prisoner's Dilemma game and for any k. Formally:

Proposition 4. *Let $E = (G_{PD}, k)$ be an environment. Defection is a robust strictly perfect equilibrium action.*

The intuition is straightforward. Consider any distribution of commitment strategies. Consider the steady state in which all the normal incumbents defect regardless of the observed signal. It is immediate that this strategy is the unique best reply to itself. This implies that if the share of committed agents is sufficiently small, then always defecting is also the unique best reply in the slightly perturbed environment.

Our first main result shows that defection is the *unique* regular perfect equilibrium in offensive games.

Theorem 1. *Let $E = (G_{PD}, k)$ be an environment, where G is an offensive Prisoner's Dilemma (i.e., $g > l$). If $(S^*, \sigma^*, \theta^*)$ is a regular perfect equilibrium, then $S^* = \{d\}$ and $\theta^* = k$.*

Sketch of Proof. The payoff of a strategy can be divided into two components: (1) a *direct* component: defecting yields additional g points if the partner cooperates and additional l points if the partner defects, and (2) an *indirect* component: the strategy's average probability of defection determines the distribution of signals observed by the partners, and thereby determines the partner's probability of defecting. For each fixed average probability of defection q the fact that the Prisoner's Dilemma is offensive implies that the optimal strategy among all those who defect with an average probability of q is to defect, with the maximal probability, against the partners who are most likely to cooperate. This implies that all agents who follow incumbent strategies are more likely to defect against partners who are more likely to cooperate. As a result, mutants who always defect outperform incumbents because they both have a strictly higher direct payoff (since defection is a dominant action) and a weakly higher indirect payoff (since incumbents are less likely to defect against them). □

Discussion of Theorem 1 The proof of Theorem 1 relies on the assumption that agents are limited to choosing only stationary strategies. The stationarity assumption implies that a partner who has been observed to defect more in the past is more likely to defect in the current match. However, this may no longer be true in a non-stationary environment. In Section 6 we analyze the classic setup of repeated games, in which agents can choose non-stationary strategies and observe the opponent's recent actions. In that setup we are able to prove a weaker version of Theorem 1 (namely, Theorem 6) which states that *full* cooperation cannot be supported as a perfect equilibrium outcome in offensive Prisoner's Dilemmas (i.e., cooperation is not a perfect equilibrium action in offensive games).

Several papers in the existing literature present various mechanisms to support cooperation in any Prisoner's Dilemma game. Kandori (1992, Theorem 1) and Ellison (1994) show that cooperation can be supported by contagious equilibria even when an agent does not observe any signal about her partner (i.e., $k = 0$). In these equilibria each agent starts the game by cooperating, but she starts defecting forever as soon as any partner has defected against her. Formally, such equilibria are impossible in our setup because the population is infinite rather than finite as in Kandori (1992) and Ellison (1994). It is possible to adapt our analysis to a setup of a large finite population (though we do not do so in this paper due to the technical difficulties) and show that the presence of a few committed agents destabilizes cooperative contagious equilibria also in large finite populations. Specifically (as pointed out by Ellison, 1994, p. 578), if we consider a large population in which at least one "crazy" agent defects with positive probability at all rounds regardless of the observed signal, then such an environment will not admit a cooperative contagious equilibrium, because agents will assign high probability to the event that the contagion process has already begun, even after having experienced a long period during which no partner defected against them.

Sugden (1986) and Kandori (1992, Theorem 2) show that cooperation can be a perfect equilibrium in a setup in which each player observes a binary signal about his partner, either a "good label" or a "bad label." All players start with a good label. This label becomes bad if a player defects against a "good" partner. The equilibrium strategy that supports full cooperation in this setup is to cooperate against good partners and defect against bad partners. Theorems 1 and 6 reveal that the presence of a small fraction of committed agents does not allow the population to maintain such a simple binary reputation under an observation structure in which players observe an arbitrary number of past actions taken by their partners. The theorem shows this indirectly, because if it were possible to derive binary reputations from this information structure, then it should have been possible to support cooperation as a perfect equilibrium action. Moreover, Theorem 4 shows that cooperation is not a perfect equilibrium action in acute games when players observe action profiles. This suggests that the presence of a few committed agents does not allow us to maintain the seemingly simple binary reputation mechanisms of Sugden (1986) and Kandori (1992), even under observation structures in which each agent observes the whole action profile of many of her opponent's past interactions.

The mild restriction to a regular perfect equilibrium is necessary for Theorem 1 to go through. Example 6 in Appendix A demonstrates the existence of a non-regular perfect equilibrium of an offensive PD, in which players cooperate with positive probability. This non-robust equilibrium is similar to the "belief-free" sequential equilibria that support cooperation in offensive Prisoner's Dilemma games in Takahashi (2010) (see also Deb, 2012), which have the property that players are always indifferent between their actions, but they choose different mixed actions depending on the signal they obtain about the partner. The example also illustrates why Takahashi's (2010) equilibria crucially depend on the absence of commitment strategies (for further discussion of the relation between our results and Takahashi, 2010, see Remark 10 below).

4.3 Stability of Cooperation in Defensive Prisoner's Dilemmas

Our second main result shows that if players observe at least two actions, then cooperation is strictly perfect in any defensive Prisoner's Dilemma. Moreover, it shows that there is essentially a unique combination of strategies that supports full cooperation in the Prisoner's Dilemma game, according to which: (a) all agents cooperate when observing no defections, (b) all agents defect when observing at least 2 defections, (3) sometimes (but not always) agents defect when observing a single defection. The average defection probability when an agent observes a single defection depends on the strategy commitments, and it is in the interval $\left[\frac{g}{l+1}\cdot\frac{1}{k},\frac{l}{l+1}\cdot\frac{1}{k}\right]$. Formally:

Theorem 2. *Let $E=(G_{PD},k)$ be an environment with observations of actions, where G_{PD} is a defensive Prisoner's Dilemma ($g<l$), and $k\geq 2$.*

1. *If $(S^*,\sigma^*,\theta^*\equiv 0)$ is a perfect equilibrium then: (a) for each $s\in S^*$, $s_0(c)=1$ and $s_m(d)=1$ for each $m\geq 2$; and (b) there exist $s,s'\in S^*$ such that $s_1(d)<1$ and $s'_1(d)>0$.*

2. *Cooperation is a robust strictly perfect equilibrium action.*

Sketch of Proof. Suppose that $(S^*,\sigma^*,\theta^*\equiv 0)$ is a perfect equilibrium. The fact that the equilibrium induces full cooperation, in the limit when the mass of commitment strategies converges to zero, implies that all normal agents must cooperate when they observe no defections, i.e., $s_0(c)=1$ for each $s\in S^*$.

Next we show that there is a normal strategy that induces the agent to defect with positive probability when observing a single defection, i.e., $s_1(d)>0$ for some $s\in S^*$. Assume to the contrary that $s_1(c)=1$ for each $s\in S^*$. If an agent (Alice) deviates and defects with small probability $\epsilon<<1$ when observing no defections,

then she outperforms the incumbents. On the one hand, the fact that she occasionally defects when observing $m = 0$ gives her a direct gain of at least $\epsilon \cdot g$. On the other hand, the probability that a partner observes her defecting twice or more is $O\left(\epsilon^2\right)$; therefore her indirect loss from these additional ϵ defections is at most $O\left(\epsilon^2\right) \cdot (1 + l)$, and therefore for a sufficiently small $\epsilon > 0$, Alice strictly outperforms the incumbents.

The fact that $s_1(d) > 0$ for some $s \in S^*$ implies that defection is a best reply conditional on an agent observing $m = 1$. The direct gain from defecting is strictly increasing in the probability that the partner defects (because the game is defensive), while the indirect influence of defection on the behavior of future partners is independent of the partner's play. This implies that defection must be the unique best reply when an agent observes $m \geq 2$ defections, since such an observation implies a higher probability that the partner is going to defect relative to the observation of a single defection. This establishes that $s_m(d) = 1$ for all $m \geq 2$ and all $s \in S^*$.

In order to demonstrate that there is a strategy s such that $s_1(d) < 1$, assume to the contrary that $s_1(d) = 1$ for each $s \in S^*$. Suppose that the average probability of defection in the population is $0 < \Pr(d)$. Since there is full cooperation in the limit we have $\Pr(d) = O(\epsilon)$. This implies that a random partner is observed to defect at least once with a probability of $k \cdot \Pr(d) + O\left(\epsilon^2\right)$. This in turn induces the defection of a fraction $k \cdot \Pr(d) + O\left(\epsilon^2\right)$ of the normal agents (under the assumption that $s_1(d) = 1$). Since the normal agents constitute a fraction $1 - O(\epsilon)$ of the population we must have $\Pr(d) = k \cdot \Pr(d) + O\left(\epsilon^2\right)$, which leads to a contradiction for any $k \geq 2$. Thus, if $s_1(d) = 1$, then defections are "contagious," and so there is no steady state in which only a fraction $O(\epsilon)$ of the population defects. This completes the sketch of the proof of part 1.

To prove part 2 of the theorem, let s^1 and s^2 be the strategies that defect iff $m \geq 1$ and $m \geq 2$, respectively. Consider the state $\left(\left\{s^1, s^2\right\}, (q^*, 1 - q^*), \theta^* \equiv 0\right)$. The direct gain from defecting (relative to cooperating) when observing $m = 1$ is

$$\Pr(m = 1) \cdot ((l \cdot \Pr(d|m = 1)) + g \cdot \Pr(c|m = 1)),$$

where $\Pr(d|m = 1)$ ($\Pr(c|m = 1)$) is the probability that a random partner is going to defect (cooperate) conditional on the agent observing $m = 1$, and $\Pr(m = 1)$ is the average probability of observing the signal $m = 1$. The indirect loss from defection, relative to cooperation, conditional on the agent observing a single defection, is

$$q^* \cdot (k \cdot \Pr(m = 1)) \cdot (l + 1) + O\left(\left(\Pr(m = 1)\right)^2\right).$$

To see this, note that a random partner defects with an average probability of q if he observes a single defection (which occurs with probability $k \cdot \Pr(m = 1)$ when the partner makes k i.i.d. observations, each of which has a probability of $\Pr(m = 1)$ of being a defection), and each defection induces a loss of $l + 1$ to the agent (who obtains $-l$ instead of 1). The fact that some normal agents cooperate and others defect when observing a single defection implies that in an equilibrium both actions have to be best replies conditional on the agent observing $m = 1$. This implies that the indirect loss from defecting is exactly equal to the direct gain (up to $O\left(\left(\Pr(m = 1)\right)^2\right)$), i.e.,

$$\Pr(m = 1) \cdot ((l \cdot \Pr(d|m = 1)) + g \cdot \Pr(c|m = 1)) = q^* \cdot (k \cdot \Pr(m = 1)) \cdot (l + 1)$$

$$\Rightarrow q^* = \frac{(l \cdot \Pr(d|m = 1)) + g \cdot \Pr(c|m = 1)}{k \cdot (l + 1)}. \tag{6}$$

The probability $\Pr(d|m = 1)$ depends on the distribution of commitments. Yet, one can show that for every distribution of commitment strategies $\left(S^C, \lambda\right)$, there is a unique value of $q^* \in (0, 1)$ that solves Eq. (6) and that, given this q^*, both s^1 and s^2 (and only these strategies) are best replies. This means that

$\left(\left\{s^1, s^2\right\}, (q^*, 1-q^*), \theta^* \equiv 0\right)$ is a perfect equilibrium. □

Discussion of Theorem 2 We comment on a few issues related to Theorem 2.

1. Each distribution of commitment strategies induces a unique frequency q^* of s^1-agents, which yields a perfect equilibrium. One may wonder whether a population starting from a different share $q_0 \neq q^*$ of s^1-agents is likely to converge to the equilibrium frequency q^*. It is possible to show that the answer is affirmative. Specifically, given any initial low frequency $q_0 \in (0, q^*)$, the s^1-agents achieve a higher payoff than the s^2-agents and, given any initial high frequency $q_0 \in \left(q^*, \frac{1}{k}\right)$, the s^1-agents achieve a lower payoff than the s^2-agents. Thus, under any smooth monotonic dynamic process in which a more successful strategy gradually becomes more frequent, the share of s^1-agents will shift from any initial value in the interval $q_0 \in \left(0, \frac{1}{k}\right)$ to the exact value of q^* that induces a perfect equilibrium.

2. As discussed in the formal proof in Appendix D.6, some distributions of commitment strategies may induce a slightly different perfect equilibrium, in which the population is homogeneous, and each agent in the population defects with probability q^* (μ) when observing a single defection (contrary to the heterogeneous deterministic behavior described above).

3. In Appendix B we show that the stability of cooperation is robust to small group of agents (with a positive small mass) who jointly deviate (à la Maynard Smith and Price's 1973 notion of evolutionary stability).

4. Our results can be extended to a setup in which the number of observed actions is random. Specifically, consider a *random environment* (G_{PD}, p), where $p \in \Delta(\mathbb{N})$ is a distribution with a finite support, and each agent privately observes k actions of the partner with probability $p(k)$. Specifically, Theorem 2 (and, similarly, Theorems 3–5) will hold for any random environment in which the probability of observing at least two interactions is sufficiently high. The perfect equilibrium has to be adapted as follows. As in the main model, all normal agents cooperate (defect) when observing no (at least two) defections. In addition, there will be a value $\bar{k} \in supp(p)$ and a probability $q \in [0,1]$ (which depend on the distribution of commitment strategies), such that all normal agents cooperate (defect) when observing a single defection out of $k > \bar{k}$ ($k < \bar{k}$), and a fraction q of the normal agents defect when observing a single defection out of $\bar{k}$ observations.

5. The threshold case between defensiveness and offensiveness: $g = l$. Such a Prisoner's Dilemma game can be interpreted as a game in which each of the players simultaneously decides whether to sacrifice a personal payoff of g in order to induce a gain of $1 + g$ to her partner. One can show that cooperation is also strictly perfect in this setup, and it is supported by the same kind of perfect equilibrium as described above. However, in this case: (1) the uniqueness result (part 1 of Theorem 2) is no longer true, as other kinds of strategies may also support full cooperation, and (2) cooperation does not satisfy the refinement of evolutionary stability (Appendix B). One can adapt the proof of Theorem 1 to show that defection is the unique perfect evolutionarily stable outcome when $g = l$.

6. *Conjecture on the convergence of the signal profile*: The robustness property proven in the formal proof of Theorem 2 implies that normal agents still cooperate with a probability very close to one also when one perturbs the profile of observed signal distributions. We *conjecture* that in future research one can show a stronger dynamical convergence result on the cooperative perfect equilibrium $\left(\left\{s^1, s^2\right\}, (q^*, 1-q^*), \theta^* \equiv 0\right)$: (1) if $k > 2$, then there is a threshold (which is decreasing in q^*) such that if the initial frequency of cooperation is smaller (larger) than this threshold, then the signal profile converges to everyone cooperating

(defecting); and (2) if $k = 2$, the signal profile converges to everyone cooperating regardless of the initial signal profile.[31]

The following example demonstrates the existence of a perfect equilibrium that supports cooperation when the unique commitment strategy is to play each action uniformly.

Example 3 (Example 1 revisited: illustration of the perfect equilibrium that supports cooperation)**.** Consider the perturbed environment $(G_D, 2, \{s^u \equiv 0.5\}, \epsilon)$, where G_D is the defensive Prisoner's Dilemma game with the parameters $g = 1$ and $l = 3$ (as presented in Table 1 in the Introduction). Consider the steady state $\left(\left\{s^1, s^2\right\}, \left(\frac{1}{6}, \frac{5}{6}\right), \theta^*\right)$, where θ^* is defined as in (4) in Example **1** above. A straightforward calculation shows that the average probability in which a normal agent observes $m = 1$ when being matched with a random partner is

$$\Pr(m = 1) = \epsilon \cdot 0.5 + 3.5 \cdot \epsilon \cdot \frac{1}{6} + 0.5 \cdot \epsilon \cdot \frac{5}{6} + O\left(\epsilon^2\right) = 1.5 \cdot \epsilon + O\left(\epsilon^2\right).$$

The probability that the partner is a committed agent conditional on observing a single defection is:

$$\Pr(s^u | m = 1) = \frac{\epsilon \cdot 0.5}{1.5 \cdot \epsilon} = \frac{1}{3} \;\Rightarrow\; \Pr(d | m = 1) = \frac{1}{3} \cdot 0.5 = \frac{1}{6},$$

which yields the conditional probability that the partner of a normal agent will defect. Next we calculate the direct gain from defecting conditional on the agent observing a single defection ($m = 1$):

$$\Pr(m = 1) \cdot ((l \cdot \Pr(d | m = 1)) + g \cdot \Pr(c | m = 1)) = 1.5 \cdot \epsilon \cdot \left(3 \cdot \frac{1}{6} + 1 \cdot \frac{5}{6}\right) + O\left(\epsilon^2\right) = 2 \cdot \epsilon + O\left(\epsilon^2\right).$$

The indirect loss from defecting conditional on the agent observing a single defection is:

$$q \cdot (k \cdot \Pr(m = 1)) \cdot (l + 1) + O\left(\epsilon^2\right) = q \cdot 2 \cdot 1.5 \cdot \epsilon \cdot (3 + 1) = 12 \cdot q \cdot \epsilon + O\left(\epsilon^2\right).$$

When taking $q = \frac{1}{6}$ the indirect loss from defecting is exactly equal to the direct gain (up to $O\left(\epsilon^2\right)$).

4.4 Stability of Cooperation when Observing a Single Action

Given a distribution of commitments $\left(S^C, \lambda\right)$, we define $\beta_{(S^C,\lambda)} \in (0, 1)$ as follows:

$$\beta_{(S^C,\lambda)} = \frac{\mathbf{E}_\lambda\left(\left(s_0(d)\right)^2\right)}{\mathbf{E}_\lambda\left(\left(s_0(d)\right)\right)} = \frac{\sum_{s \in S^C} \lambda(s) \cdot \left(s_0(d)\right)^2}{\sum_{s \in S^C} \lambda(s) \cdot s_0(d)}. \tag{7}$$

The value of $\beta_{(S^C,\lambda)}$ is the ratio between the mean of the square of the probability of defection of a random committed agent who observes $m = 0$ and the mean of the same probability without squaring it. In particular, when the set of commitments is a singleton, $\beta_{(S^C,\lambda)}$ is equal to the probability that a committed agent defects when she observes $m = 0$ (i.e., $\beta_{(S^C,\lambda)} = s_0(d)$).

The following result shows that if the game is defensive and agents observe a single action, then full cooperation is a perfect equilibrium action with respect to the distribution of commitments $\left(S^C, \lambda\right)$ iff $g \leq \beta_{(S^C,\lambda)}$. In particular, cooperation is a (non-strictly) perfect equilibrium action iff $g < 1$.

[31] The intuition for this global convergence conjecture is that if $k = 2$ and $k \cdot q < 1$, then cooperation is "contagious", and a small number of committed agents who cooperate is sufficient to move the entire population towards a state where everyone cooperates (due to an an analogous argument, presented in the proof of Theorem 3, which explains why defections are contagious in acute Prisoner's Dilemmas).

Proposition 5. *Let $E = (G_{PD}, 1)$ be an environment, where G_{PD} is a defensive Prisoner's Dilemma ($g < l$). Let $\left(S^C, \lambda\right)$ be a distribution of commitments. There exists a perfect equilibrium $(S^*, \sigma^*, \theta^* \equiv 0)$ with respect to $\left(S^C, \lambda\right)$ iff $g \leq \beta_{(S^C,\lambda)}$.*

Sketch of Proof. Similar arguments to those presented in part 1 of Theorem 2 imply that any distribution of commitment strategies induces a unique average probability q by which normal agents defect when observing $m = 1$, in any cooperative perfect equilibrium. This implies that a deviator who always defects gets a payoff of $1 + g$ in a fraction $1 - q$ of the interactions. One can show that such a deviator outperforms the incumbents iff[32] $g > \beta_{(S^C,\lambda)}$. □

Corollary 1. *Let $E = (G_{PD}, 1)$ be an environment, where G_{PD} is a defensive Prisoner's Dilemma ($g < l$). Cooperation is a perfect equilibrium action iff $g < 1$.*

The intuition for the difference between the case of $k = 1$ and the case of $k \geq 2$ is that the higher g is, the more severely defection has to be punished. However, since $k = 1$, defection can be deterred only by increasing the probability of an agent defecting upon observing $m = 1$, and hence this deterrent is not enough if g is sufficiently high.

5 General Observation Structures

In this section we extend our analysis to general observation structures in which the signal about the partner may also depend on the behavior of other opponents against the partner.

5.1 Definitions

An *observation structure* is a tuple $\Theta = (k, B, o)$, where $k \in \mathbb{N}$ is the number of observed interactions, $B = \left\{b_1, .., b_{|B|}\right\}$ is a finite set of *observations* that can be made in each interaction, and the mapping $o : A \times A \to \Delta(B)$ describes the probability of observing each signal $b \in B$ conditional on the action profile played in this interaction (where the first action is the one played by the current partner, and the second action is the one played by her opponent). Note that observing actions (which was analyzed in the previous section) is equivalent to having $B = A$ and $o(a, a') = a$.

In the results of this section we focus on three observation structures:

1. *Observation of action profiles:* $B = A^2$ and $o(a, a') = (a, a')$. In this observation structure, each agent observes, in each sampled interaction of her partner, both the action played by her partner and the action played by her partner's opponent.

2. *Observation of conflicts* (in PDs): observing whether or not there was mutual cooperation. That is, $B = \{C, D\}$, $o(c, c) = C$, and $o(a, a') = D$ for any $(a, a') \neq (c, c)$. Such an observation structure (which we have not seen in the existing literature) seems like a plausible way to capture non-verifiable feedback about the partner's behavior. The agent can observe, in each sampled past interaction of the partner, whether both partners were "happy" (i.e., mutual cooperation) or whether the partners complained about each other (i.e., there was a conflict, at least one of the players defected, and it is too costly for an outside observer to verify who actually defected).

[32] In environments with $k \geq 2$, a deviator who always defects gets a payoff of zero, regardless of the value of q (because all agents observe $m = k$ when being matched with such a deviator).

3. Observation of actions against cooperation: $B = \{CC, DC, *D\}$ and $o(c,c) = CC$, $o(d,c) = DC$, and $o(c,d) = o(d,d) = *D$. That is, each agent (Alice) observes a ternary signal about each sampled interaction of her partner (Bob): either both players cooperated, or Bob unilaterally defected, or Bob's partner defected (and in this latter case Alice cannot observe Bob's action). We analyze this observation structure because it turns out to be an "optimal" observation structure that allows cooperation to be supported as a perfect equilibrium action in any Prisoner's Dilemma.

In each of these cases, we let the mapping o and the set of signals B be implied by the context, and identify the observation structure Θ with the number of observed interactions k.

In what follows we present the definitions of the main model (Sections 2 and 3) that have to be changed to deal with the general observation structure. Before playing the game, each player independently samples k independent interactions of her partner. Let M denote the set of feasible signals:

$$M = \left\{ m \in \mathbb{N}^{|B|} \middle| \sum_i m_i = k \right\},$$

where m_i is interpreted as the number of times that observation b_i has been observed in the sample. When the underlying game is the Prisoner's Dilemma and agents observe conflicts, we simplify the notation by letting $M = \{1, ..., k\}$, and interpreting $m \in \{1, ..., k\}$ as the number of observed conflicts.

The definitions of a strategy and a perturbed environment remain the same. Given a distribution of action profiles $\psi \in \triangle(A \times A)$, let $\nu_\psi = \nu(\psi) \in \Delta(M)$ be the multinomial distribution of signals that is induced by the distribution of action profiles ψ, i.e.,

$$\nu_\psi\left(m_1, ..., m_{|B|}\right) = \frac{k!}{m_1! \cdot ... \cdot m_{|B|}!} \cdot \prod_{i=1}^{|B|} \left(\sum_{(a,a') \in A \times A} \left(\psi(a,a') \cdot \left(o(a,a')(b_i) \right) \right) \right)^{m_i}.$$

The definition of a steady state is adapted as follows.

Definition 9 (Adaptation of Def. 4)**.** A *steady state* (or *state*) of a perturbed environment $\left((G,k), \left(S^C, \lambda\right), \epsilon\right)$ is a triple (S, σ, θ), where $S \subseteq \mathcal{S}$ is a finite set of strategies, $\sigma \in \Delta(S)$ is a distribution, and $\theta : \left(S \cup S^C\right) \to \triangle(M)$ is a profile of signal distributions that satisfies for each signal m and each strategy s the consistency requirement (9) below. Let $\psi_s \in \triangle(A \times A)$ be the (possibly correlated) distribution of action profiles that is played when an agent with strategy $s \in S \cup S^C$ is matched with a random partner (given σ and θ); i.e., for each $(a, a') \in A \times A$, where a is interpreted as the action of the agent with strategy s, and a' is interpreted as the action of her partner, let exists

$$\psi_s(a,a') = \sum_{s' \in S \cup S^C} \left((1-\epsilon) \cdot \sigma(s') + \epsilon \cdot \lambda(s') \right) \cdot s(\theta_{s'})(a) \cdot s'(\theta_s)(a'). \tag{8}$$

The *consistency requirement* that the mapping θ has to satisfy is

$$\forall m \in M,\ s \in S \cup S^C,\quad \theta_s(m) = \nu(\psi_s)(m). \tag{9}$$

The definition of the long-run payoff of an incumbent agent remains unchanged. We now adapt the definition of the payoff of an agent (Alice) who deviates and plays a non-incumbent strategy. Unlike in the basic model, in this extension there might be multiple consistent outcomes following Alice's deviation, as demonstrated in Example 4.

Example 4. Consider an unperturbed environment $(G_{PD}, 3)$ with an observation of $k = 3$ action profiles. Consider a homogeneous incumbent population in which all agents play the following strategy: $s^*(m) = d$ if m includes at least 2 interactions with (d, d), and $s^*(m) = c$ otherwise. Consider the state $(\{s^*\}, \theta^* = 0)$ in which everyone cooperates. Consider a deviator (Alice) who follows the strategy of always defecting. Then there exist three consistent post-deviation steady states (in all of which the incumbents continue to cooperate among themselves): (1) all the incumbents defect against Alice, (2) all the incumbents cooperate against Alice, and (3) all the incumbents defect against Alice with a probability of 50%.

Formally, we define a consistent distribution of signals for a deviator as follows.

Definition 10. Given steady state (S, σ, θ) and non-incumbent strategy $\hat{s} \in \mathcal{S} \setminus \left(S \cup S^C\right)$, we say that a distribution of signals $\theta_{\hat{s}} \in \Delta(M)$ is *consistent* if

$$\forall m \in M, \quad \theta_{\hat{s}}(m) = \nu\left(\psi_{\hat{s}}\right)(m),$$

where $\psi_s \in \triangle(A \times A)$ is defined as in (8) above. Let $\Theta_{\hat{s}} \subseteq \Delta(M)$ be the set of all *consistent signal distributions* of strategy $\hat{s}$.

Given steady state (S, σ, θ), non-incumbent strategy $\hat{s} \in \mathcal{S} \setminus \left(S \cup S^C\right)$, and consistent signal distribution $\theta(s) \equiv \theta_{\hat{s}} \in \Delta(M)$, let $\pi_{\hat{s}}(S, \sigma, \theta | \theta_{\hat{s}})$ denote the deviator's (long-run) payoff given that in the post-deviation steady state the deviator's distribution of signals is $\theta_{\hat{s}}$. Formally:

$$\pi_{\hat{s}}(S, \sigma, \theta | \theta_{\hat{s}}) = \sum_{s' \in S \cup S^C} \left((1-\epsilon) \cdot \sigma(s') + \epsilon \cdot \lambda(s')\right) \cdot \left(\sum_{(a,a') \in A \times A} \hat{s}_{\theta(s')}(a) \cdot s'_{\theta(\hat{s})}(a') \cdot \pi(a, a') \right).$$

Let $\pi_{\hat{s}}(S, \sigma, \theta)$ be the maximal (long-run) payoff for a deviator who follows strategy $\hat{s}$ in a post-deviation steady state:

$$\pi_{\hat{s}}(S, \sigma, \theta) :=_{\theta_{\hat{s}} \in \Theta_{\hat{s}}} \max \pi_{\hat{s}}(S, \sigma, \theta | \theta_{\hat{s}}). \tag{10}$$

Remark 6. Our results remain the same if one replaces the maximum function in (10) with a minimum function.

5.2 Acute and Mild Prisoner's Dilemma

In this subsection we present a novel classification of Prisoner's Dilemma games that plays an important role in the results of this section. Recall that the parameter g of a Prisoner's Dilemma game may take any value in the interval $[0, l+1]$ (if $g > l+1$, then mutual cooperation is no longer the efficient outcome that maximizes the sum of payoffs). We say that a Prisoner's Dilemma game is *acute* if g is in the upper half of this interval (i.e., if $g > \frac{l+1}{2}$), and *mild* if it's in the lower half (i.e., if $g < \frac{l+1}{2}$). The threshold, $g = \frac{l+1}{2}$, is characterized by the fact that the gain from a single unilateral defection is exactly half the loss incurred by the partner who is the sole cooperator. Hence, unilateral defection is *mildly tempting* in mild games and *acutely tempting* in acute games. An interpretation of this threshold comes from a setup (which will be important for our results) in which an agent is deterred from unilaterally defecting because it induces future partners to unilaterally defect against the agent with some probability. Deterrence in acute games requires this probability of being punished to be more than 50%, while a probability of below 50% is enough for mild games.

Example 5. Table 4 demonstrates the payoffs of specific acute (G_A) and mild (G_M) Prisoner's Dilemma games. In both examples $g = l$, i.e., the Prisoner's Dilemma game is "linear." This means that it can be

described as a "helping game" in which agents have to decide simultaneously whether to give up a payoff of g in order to create a benefit of $1+g$ for the partner. In the acute game (G_A) on the left, $g=3$ and the loss of a helping player amounts to more than half of of the benefit to the partner who receives the help ($\frac{3}{3+1}=\frac{3}{4}>\frac{1}{2}$), while in the mild game (G_M) on the right, $g=0.2$ and the loss of the helping player is less than half of the benefit to the partner who receives the help ($\frac{0.2}{0.2+1}=\frac{1}{6}<\frac{1}{2}$).

Table 4: Matrix Payoffs of Acute and Mild Prisoner's Dilemma Games

	c	d
c	1 \ 1	$1+g$ \ $-l$
d	$-l$ \ $1+g$	0 \ 0

General Prisoner's Dilemma
G_{PD}: $g,l>0$, $g<l+1$

	c	d
c	1 \ 1	4 \ -3
d	-3 \ 4	0 \ 0

Ex. 3: Acute Prisoner's Dilemma
G_A: $g=l=3>\frac{l+1}{2}=2$

	c	d
c	1 \ 1	1.2 \ -0.2
d	-0.2 \ 1.2	0 \ 0

Ex. 4: Mild Prisoner's Dilemma
G_M: $g=l=0.2<\frac{l+1}{2}=0.6$

5.3 Analysis of the Stability of Cooperation

We first note that Proposition 4 is valid also in this extended setup, with minor adaptations to the proof. Thus, always defecting is a strictly perfect equilibrium regardless of the observation structure. Next we analyze the stability of cooperation in each of the three interesting observation structures.

The following two results show that under either **observation of conflicts** or **observation of action profiles**, cooperation is a perfect equilibrium iff the Prisoner's Dilemma is mild. Moreover, in mild Prisoner's Dilemma games there is essentially a unique strategy distribution that supports cooperation (which is analogous to the essentially unique strategy distribution in Theorem 2). Formally:

Theorem 3. *Let $E=(G,k)$ be an environment with observation of conflicts, where G is a PD and $k\geq 2$.*

1. *If G is a mild PD ($g<\frac{l+1}{2}$), then:*
   (a) *If $(S^*,\sigma^*,\theta^*\equiv 0)$ is a perfect equilibrium then (1) for each $s\in S^*$, $s_0(c)=1$ and $s_m(d)=1$ for each $m\geq 2$, and (2) there exist $s,s'\in S^*$ such that $s_1(d)<1$ and $s'_1(d)>0$.*
   (b) *Cooperation is a strictly perfect equilibrium action.*
2. *If G is an acute PD ($g>\frac{l+1}{2}$), then cooperation is not a perfect equilibrium action.*

Sketch of proof. The argument for part 1(a) is analogous to Theorem 2. In what follows we sketch the proofs of part 1(b) and part 2. Fix a distribution of commitments, and a commitment level $\epsilon\in(0,1)$. Let m denote the number of observed conflicts and define s^1 and s^2 as before, but with the new meaning of m. Consider the following candidate for a perfect equilibrium $\left(\left\{s^1,s^2\right\},(q,1-q),\theta^*\equiv 0\right)$. Here, the probability q will be determined such that both actions are best replies when an agent observes a single conflict. That is, the direct benefit from her defecting when observing $m=1$ (the LHS of the equation below) must balance the indirect loss due to inducing future partners who observe these conflicts to defect (the RHS, neglecting terms of $O(\epsilon)$). The RHS is calculated by noting that defection induces an additional conflict only if the current partner has cooperated and that, on expectation, each such additional conflict is observed by k future partners, each of

whom defects with an average probability of q). Recall that $\Pr(d|m=1)$ ($\Pr(c|m=1)$) is the probability that a random partner is going to defect (cooperate) conditional on the agent observing $m=1$.

$$\Pr(m=1)\cdot((l\cdot\Pr(d|m=1))+g\cdot\Pr(c|m=1))=\Pr(m=1)\cdot k\cdot q\cdot\Pr(c|m=1)\cdot(l+1)$$
$$\Leftrightarrow q\cdot k=\frac{(l\cdot\Pr(d|m=1))+g\cdot\Pr(c|m=1)}{\Pr(c|m=1)\cdot(l+1)}. \tag{11}$$

One can see that the RHS is increasing in $\Pr(d|m=1)$. The minimal bound on the value of q is obtained when $\Pr(d|m=1)=0$. In this case $q\cdot k=\frac{g}{l+1}$.

Suppose that the game is acute. In this case $q\cdot k>0.5$. Suppose that the average probability of defection in the population is $\Pr(d)$. Since there is full cooperation in the limit we have $\Pr(d)=O(\epsilon)$. This implies that a fraction $2\cdot\Pr(d)+O(\epsilon^2)$ of the population is involved in conflicts. This in turn induces the defection of a fraction $2\cdot\Pr(d)\cdot k\cdot q+O(\epsilon^2)$ of the normal agents (because a normal agent defects with probability q upon observing at least one conflict in the k sampled interactions). Since the normal agents constitute a fraction $1-O(\epsilon)$ of the population we must have $\Pr(d)=2\cdot\Pr(d)\cdot k\cdot q+O(\epsilon^2)$. However, in an acute game, $2\cdot k\cdot q>1$ leads to the contradiction that $\Pr(d)<\Pr(d)$. Thus, if $2\cdot k\cdot q>1$, then defections are contagious, and so there is no steady state in which only a fraction $O(\epsilon)$ of the population defects.

Suppose that the game is mild. One can show that $\Pr(d|m=1)$ is decreasing in q, and that it converges to zero when $k\cdot q\nearrow 0.5$. (The reason is that when $k\cdot q$ is close to 0.5 each defection by a committed agent induces many defections by normal agents and, conditional on observing $m=1$, the partner is likely to be normal and to cooperate when being matched with a normal agent.) It follows that the RHS of Eq. (11) is decreasing in q and approaches the value $\frac{g}{l+1}$ when $k\cdot q\nearrow 0.5$. Since the game is mild, $\frac{g}{l+1}<0.5$. Hence there is some $q\cdot k<0.5$ that solves Eq. (11), and in which the normal agents defect with a low probability of ($O(\epsilon)$). □

Theorem 4. *Let $E=(G_{PD},k)$ be an environment with observation of action profiles and $k\geq 2$.*

1. *If G is a mild PD ($g<\frac{l+1}{2}$), then cooperation is a perfect equilibrium action.*
2. *If G is an acute PD ($g>\frac{l+1}{2}$), then cooperation is not a perfect equilibrium action.*

Sketch of proof. Using arguments that are familiar from above one can show that in any perfect equilibrium that supports cooperation, normal agents have to defect with an average probability of $q\in(0,1)$ when observing a single unilateral defection (and $k-1$ mutual cooperations), and defect with a smaller probability when observing a single mutual defection (since this is necessary in order for a normal agent to have better incentives to cooperate against a partner who is more likely to cooperate). The value of q is determined by Eq. (11) above, implying that both actions are best replies conditional on an agent observing the partner to be the sole defector once, and to be involved in mutual cooperation in the remaining $k-1$ observed action profiles. Let ϵ be the share of committed agents, and let φ be the average probability that a committed agent unilaterally defects. In order to simplify the sketch of the proof, we will focus on the case in which the committed agents defect with a small probability when observing the partner to have been involved only in mutual cooperations, which implies, in particular, that $\varphi<<1$ (the formal proof in the Appendix does not make this simplifying assumption). The unilateral defections of the committed agents induce a fraction $\epsilon\cdot\varphi\cdot k\cdot q+O(\epsilon^2)+O(\varphi^2)$ of the normal agents to defect when being matched against committed agents (because a normal agent defects with probability q upon observing a single unilateral defection in the k sampled interactions). These unilateral defections of normal agents against committed agents induce a further $(\epsilon\cdot\varphi\cdot k\cdot q)\cdot k\cdot q+O(\epsilon^2)$ defections of normal agents against other normal agents. Repeating this argument we come to the conclusion that the

average probability of a normal agent being the sole defector is (neglecting terms of $O\left(\epsilon^2\right)$ and $O\left(\varphi^2\right)$):

$$\epsilon \cdot \varphi \cdot k \cdot q \cdot \left(1 + k \cdot q + (k \cdot q)^2 + ...\right) = \epsilon \cdot \varphi \cdot \frac{k \cdot q}{1 - k \cdot q.}$$

As discussed above, in acute games, the value of $k \cdot q$ must be larger than 0.5, which implies that $\frac{k \cdot q}{1-k \cdot q} > 1$. This implies that conditional on an agent observing the partner to be the sole defector once, the posterior probability that the partner is normal is:

$$\frac{\epsilon \cdot \varphi \cdot \frac{k \cdot q}{1-k \cdot q}}{\epsilon \cdot \varphi + \epsilon \cdot \varphi \cdot \frac{k \cdot q}{1-k \cdot q}} = \frac{\frac{k \cdot q}{1-k \cdot q}}{1 + \frac{k \cdot q}{1-k \cdot q}} > 0.5.$$

Thus, normal agents are more likely to unilaterally defect than committed agents. One can show that when there is a mutual defection, it is most likely that at least one of the agents involved is committed. This implies that the partner is more likely to defect when he is observed to be involved in mutual defection relative to being observed to be the sole defector. This implies that defection is the unique best reply when observing a single mutual defection, and this contradicts the assumption that normal agents cooperate with positive probability when observing a single mutual defection. When the game is mild, a construction similar to the previous proofs supports cooperation as a perfect equilibrium. □

Our last result studies the observation of actions against cooperation, and it shows that cooperation is a perfect equilibrium action in any underlying Prisoner's Dilemma. Formally:

Theorem 5. *Let $E = (G, k)$ be an environment with observation of actions against cooperation, where G is a PD game and $p \equiv k \geq 2$. Then cooperation is a perfect equilibrium action.*

The intuition behind the proof is as follows. Not allowing Alice to observe Bob's behavior when his past opponent has defected helps to sustain cooperation because it implies that defecting against a defector does not have any negative indirect effect (in any steady state) because it is never observed by future opponents. This encourages agents to defect against partners who are more likely to defect, and allows cooperation to be sustained regardless of the values of g and l.

Remark 7. In the last two results (Theorems 4 and 5) cooperation is not a strictly perfect equilibrium. Specifically, it is not a perfect equilibrium action with respect to distributions of commitments in which the committed agents defect with high probability. The reason is that committed agents who defect with high probability induce normal partners to defect against them with probability one. This implies that when observing a partner to be involved on either side of a unilateral defection (either as the sole defector or as the sole cooperator), the partner is most likely to be normal. As a result the agents' incentives to defect are the same when observing mutual cooperation as when observing unilateral defection, and this does not allow cooperation to be supported in a perfect equilibrium, as such cooperation relies on agents who have better incentives to defect when observing a unilateral defection.

6 Conventional Repeated Game Model

The main model of the paper relies on various simplifying assumptions, and some unconventional modeling choices that distinguish it from the existing literature: (1) the interactions within the community do not have an explicit starting point, (2) agents live forever and do not discount the future, (3) agents are only allowed to follow stationary strategies, (4) agents observe the partner's actions sampled from the entire infinite history

of play of the partner, and (5) an agent does not know the actions observed by her partner about the agent's behavior. In this section we present a conventional repeated game model that relaxes all of these assumptions. It differs from most of the existing literature in only one respect: the presence of a small fraction of committed agents in the population. We show that this difference is sufficient to yield most of our key results.

For brevity, we focus only on the observations of actions. The adaptation of the results on general observation structures is analogous.

6.1 Adaptations to the Model

Environment as a Repeated Game We consider an infinite population (a continuum of mass one) interacting in discrete time $t = 0, 1, 2, 3, ...$ We redefine an *environment* to be a triple (G, k, δ), where $G = (A, \pi)$ is the underlying symmetric game, $k \in \mathbb{N}$ is the number of recent actions of an agent that are observed by her partner, and $\delta \in (0, 1)$ is the discount factor of the agents. In each period the agents are randomly matched into pairs and, before playing, each agent observes the most recent $\min(k, t)$ actions of her partner; i.e., an agent observes all past actions in the early rounds when $t \leq k$, and she observes only the last k rounds in later rounds when $t > k$. Let $M = \cup_{i \leq k} A^i$ denote the set of all possible signals.

Remark 8. Our assumption that each agent observes the last k actions of the partner is made only to simplify the notation and to allow for a more concise appendix. Our results can be adapted to a more general setup in which each agent observes k actions randomly sampled from the partner's last $n \geq k$ actions. The case of $n >> k$ is the one closest to the main model. We choose to focus on the opposite case of $n = k$ (i.e., observation of the last k actions) in order to demonstrate the robustness of our results in the setup that is the "furthest" from the main model.

A (private) history of an agent at round $\hat{t}$ is a tuple $h_{\hat{t}} = \left((m_t, a_t, b_t)_{0 \leq t < \hat{t}}, m_{\hat{t}}\right)$, where $m_t \in M$ is the signal observed by the agent about her opponent at round t, $a_t \in A$ is the action played by the agent at round t, and $b_t \in A$ is the action played by the past partner at round t. Finally, $m_{\hat{t}}$ is the signal the agent has observed about her current partner. Let $H_{\hat{t}}$ denote the set of all possible histories at round $\hat{t}$, and let $H = \cup_{T \in \mathbb{N}} H_T$ denote the set of all (finite) histories.

A *strategy* is a mapping $s : H \to \Delta(A)$ assigning a mixed action to each (private) history. We redefine S to denote the set of all such strategies. Note, that unlike in the main model we do not impose any restrictions on the set of feasible strategies. In particular, we allow agents to follow non-stationary strategies. A strategy is *uniformly totally mixed* if there exist $\gamma > 0$ such that for each history $h_{\hat{t}} \in H$ and each action $a \in A$ it is the case that $s_{h_{\hat{t}}}(a) > \gamma$.

Perturbed Environment and Population State A perturbed environment is a tuple consisting of: (1) an environment, (2) a distribution λ over a set of commitment strategies S^C that includes a uniformly totally mixed strategy, and (3) a number ϵ representing how many agents are committed to playing strategies in S^C (*committed agents*). The remaining $1 - \epsilon$ agents can play any strategy in S^N (*normal agents*). Formally:

Definition 11. A *perturbed environment* is a tuple $E_\epsilon = \left((G, k, \delta), \left(S^C, \lambda\right), \epsilon\right)$, where (G, k, δ) is an environment, S^C is a non-empty finite set of strategies (called *commitment strategies*) that includes a uniformly totally mixed strategy, $\lambda \in \Delta\left(S^C\right)$ is a distribution with full support over the commitment strategies, and $\epsilon \geq 0$ is the mass of committed agents in the population.

A *population state* is defined as a pair $\left(S^N, \sigma\right)$, where S^N is the finite set of normal strategies in the population, and $\sigma \in \Delta\left(S^N\right)$ is the distribution describing the frequency of each normal strategy in the population

of normal agents. By standard arguments, a population state $\left(S^N, \sigma\right)$ and a perturbed environment E_ϵ jointly induce a unique sequence of distributions over the set of histories. Formally, there exists a unique profile $\left(\mu_{s,t}\right)_{s\in S, t\in\mathbb{N}}$, where each $\mu_{s,t} \in \Delta\left(H_T\right)$ is a distribution over the histories of length t, such that $\mu_{s,t}\left(h_t\right)$ is the probability that an agent who follows strategy s reaches history $h_t \in H_t$ in round t.

Expected Payoff and Equilibria In what follows we define the (ex-ante) expected payoff of an agent who follows strategy s and has discount factor δ , given a population state $\left(S^N, \sigma\right)$ of a perturbed environment $E_\epsilon = \left(\left(G, k, \delta\right), \left(S^C, \lambda\right), \epsilon\right)$. When $s \in S^N \cup S^C$ is an incumbent strategy, we define the payoff as follows:

$$\pi_s\left(S^N, \sigma, E_\epsilon\right) = \left(1-\delta\right) \cdot \sum_{t\geq 1} \delta^{t-1} \cdot \sum_{h_t = \left((m_t, a_t, b_t)_{0\leq t<\hat{t}}, m_{\hat{t}}\right) \in H^t} \mu_{s,t}\left(h_t\right) \cdot \pi\left(a_{t-1}, b_{t-1}\right). \tag{12}$$

As in the stationary model, let $\pi\left(S^N, \sigma, E_\epsilon\right) = \sum_{s\in S^N} \sigma\left(s\right) \cdot \pi_s\left(S^N, \sigma, E_\epsilon\right)$ denote the mean payoff of the normal agents in the population.

Next consider an agent (Alice) who deviates and plays a new strategy $\hat{s} \in \mathcal{S} \backslash S^N$. Alice's strategy determines her behavior against the incumbents. This determines the distribution of signals that are observed by the partners when being matched with Alice, and thus it determines the incumbents' play against Alice, and this uniquely determines the sequence of distributions over the set of histories of Alice. Formally, there exists a unique profile $\left(\mu_{\hat{s},t}\right)_{t\in\mathbb{N}}$, where each $\mu_{\hat{s},t} \in \Delta\left(H_T\right)$ is a distribution over the histories of length t, such that $\mu_{\hat{s},t}\left(h_t\right)$ is the probability that Alice who follows strategy $\hat{s}$ reaches history $h_t \in H_t$ in round t. We define Alice's payoff $\pi_{\hat{s}}\left(S^N, \sigma, E_\epsilon\right)$ in the same way as in Eq. (12), with $\mu_{\hat{s},t}\left(h_t\right)$ replacing $\mu_{s,t}\left(h_t\right)$.

The definition of Nash equilibrium is standard:

Definition 12. A population state $\left(S^N, \sigma\right)$ of the perturbed environment $\left(\left(G, k, \delta\right), \left(S^C, \lambda\right), \epsilon\right)$ is a *Nash equilibrium* if for each strategy $s \in \mathcal{S}$, it is the case that $\pi_s\left(S^N, \sigma, E_\epsilon\right) \leq \pi\left(S^N, \sigma, E_\epsilon\right)$.

Remark 9. By standard arguments one can show that the set of *Nash* equilibrium payoffs in a perturbed environment coincides with the set of *sequential* equilibrium payoffs, due to the fact that every history of play about the partners is observed with a positive probability due to the presence of totally mixed commitment strategies in the population.

Definition 13. Fix an environment (G, k, δ). A sequence of strategies $\left(s_n\right)_n$ converges to strategy s (denoted by $\left(s_n\right)_n \to_{n\to\infty} s$) if for each round $t \in \mathbb{N}$, each history $h_t \in H_t$, and each action a, the sequence of probabilities $s\left(h_t\right)\left(a\right)$ converges to $s\left(h_t\right)\left(a\right)$. A sequence of population states $\left(S^N{}_n, \sigma_n\right)_n$ converges to a population state $\left(S^N, \sigma^*\right)$ if for each strategy $s \in supp\left(\sigma^*\right)$, there exists a sequence of sets of strategies $\left(S_n^N\right)_n$ such that: (1) $\sum_{s_n \in S_n^N} \sigma_n\left(s_n\right) \to \sigma^*\left(s\right)$, and (2) for each sequence of elements of those sets (i.e., for each sequence of strategies $\left(s_n\right)_n$ such that $s_n \in S_n^N$ for each n), $s_n \to_{n\to\infty} s$.

A perfect equilibrium is defined as the limit of a converging sequence of Nash equilibria of a converging sequence of perturbed environments. Formally:

Definition 14. A population state $\left(S^N, \sigma^*\right)$ of the environment (G, k, δ) is a *perfect equilibrium* if there exist a distribution of commitments $\left(S^C, \lambda\right)$ and converging sequences $\left(S_n^N, \sigma_n\right)_n \to_{n\to\infty} \left(S^N, \sigma^*\right)$ and $\left(\epsilon_n > 0\right)_n \to_{n\to\infty} 0$, such that for each n, the state $\left(S_n^N, \sigma_n\right)$ is a Nash equilibrium of the perturbed environment $\left(\left(G, k, \delta\right), \left(S^C, \lambda\right), \epsilon_n\right)$. If the underlying game is the Prisoner's Dilemma, we say that the *perfect equilibrium induces full cooperation* if $\lim_{n\to\infty} \pi\left(S_n^N, \sigma, E_{\epsilon_n}\right) = \pi\left(c, c\right)$.

We say that cooperation is a perfect equilibrium outcome if there exists a perfect equilibrium that induces full cooperation.

6.2 Adaptation of Main Results

The following result adapts the main results of Section 4. Specifically, it shows that full cooperation is a perfect equilibrium outcome iff the underlying Prisoner's Dilemma game is (weakly) defensive. Moreover, the way in which full cooperation is sustained in defensive games is similar to the strategy presented in the stationary model, although it includes an additional technical complication due to the non-stationarity. The intuition for the result is similar to the intuition described in connection with the results of the stationary model.

Theorem 6. *Let (G_{PD}, k, δ) be an environment with a Prisoner's Dilemma underlying the game and $k \geq 2$.*

1. *Cooperation is not a perfect equilibrium outcome if $g > l$.*
2. *Cooperation is a perfect equilibrium outcome if $g \leq l$ and $\delta^k > \frac{l}{l+1}$. Moreover, cooperation is sustained by a strategy in which each normal agent (1) always cooperates if she observes the partner always cooperating, (2) always defects if she observes the partner defecting at least twice, and (3) sometime defects if she observes the partner defecting once.*

6.3 Discussion of the Results in the Setup of Repeated Games

Theorem 6 adapts our main results from the stationary model (Theorems 1 and 2) to the conventional setup of repeated games. The adaptation weakens our main results in three aspects:

1. While Theorem 1 shows that no level of partial cooperation is sustainable in stationary environments, Theorem 6 shows only that full cooperation is not sustainable. The reason for this is as follows. In stationary environments, if the partner has been observed to defect more often in the past it is implies that he is more likely to defect in the current match. Such an inference is much more complicated in a non-stationary environment in which committed agents may behave in a complicated non-stationary way. In such setups, we are able to make only a weaker inference: if all normal agents almost always cooperate, then if an agent observes the partner to defect in the past, it increases the probability that the partner will defect in the current match (relative to observing the partner always cooperating).

2. While Theorem 2 shows that there is essentially a unique way to support full cooperation, Theorem 1 shows only that a somewhat more complicated version of the same mechanism can also be used to support full cooperation in standard repeated games. The fact that we allow non-stationary strategies and that observed actions are ordered induces a much larger set of strategies, and does not allow us to show a similar uniqueness property in this setup.

3. Theorem 2 shows that the cooperation is a *strictly* perfect equilibrium outcome; i.e., cooperation can be sustained regardless of the behavior of the committed agents. In this setup, as there is a much larger set of non-stationary strategies that may be used by committed agents, we are not able to show a similar strictness property.

Remark 10 (Comparison with Takahashi, 2010). The setup in this section is almost identical to the setup of Takahashi (2010). Specifically, the only key difference between the two models is that we introduce the presence of a few committed agents to the population (in addition, Takahashi deals with a setup in which

an agent observes all past actions of the partner, but one can adapt his results to a setup in which an agent observes only the recent k actions of the partner). Takahashi (2010, Prop. 2) constructs "belief-free" sequential equilibria in which (1) each agent is indifferent between the two actions after any history, and (2) each agent chooses actions independent of her own record of past play. Takahashi shows how these equilibria can support any level of cooperation, and, in particular, can support full cooperation in any Prisoner's Dilemma.

The key contribution of this section is showing that the presence of few a committed agents (regardless of the commitment strategies they follow) substantially changes this result. When committed agents are present, an agent can no longer be indifferent between the two actions after all histories of play, and can no longer play in the current match independently of her own record of past play. We adapt Takahashi's construction and present an equilibrium in which each agent is indifferent between the two actions after only one class of histories: those in which the agent has cooperated in the previous $k-1$ rounds and she observes the partner to defect only in the last round. In all other classes of histories, the agents have strict incentives to either cooperate or defect. In defensive games, the strict incentives are in the "good" direction that allows one to sustain full cooperation. In offensive games, the strict incentives are in the opposite direction, and one cannot sustain full cooperation.

Elsewhere, Heller (2017) presents a more general non-robustness argument in the standard setup of repeated games played by the *same* two players, and shows that none of the belief-free equilibria are robust against small perturbations in the behavior of potential opponents (in the sense of not satisfying a very mild refinement in the spirit of evolutionary stability).

Takahashi (2010) also shows that any defensive Prisoner's Dilemma admits a strict equilibrium that induces full cooperation in defensive games; this equilibrium relies on the agents observing many past actions and on following a relatively complex strategy.[33] Our result shows that the presence of a few committed agents allow agents to sustain stable cooperation in defensive Prisoner's Dilemma games by using simple strategies that rely on agents observing only two of the partner's past actions.

7 Discussion

7.1 Related Literature

In what follows we discuss related literature that has not been discussed elsewhere in the paper.

Image Scoring In an influential paper, Nowak and Sigmund (1998) presents the mechanism of *image scoring* to support cooperation when agents from a large community are randomly matched and each agent observes the partner's past actions. In their setup, each agent observes the last k past actions of the partner, and she defects if and only if the partner has defected at least m times in the last k observed actions. A couple of papers have raised concerns about the stability of cooperation under image-scoring mechanisms. Specifically, Leimar and Hammerstein (2001) demonstrate in simulations that cooperation is unstable, and Panchanathan and Boyd (2003) analytically study the case in which each agent observes the last action.[34] Our paper makes two key contributions to this literature. First, we introduce a novel variant of image scoring that is essentially the unique stationary way to support cooperation as a perfect equilibrium outcome when agents observe actions. Second, we show that the classification of Prisoner's Dilemma games into offensive and

[33] The strategy in Takahashi (2010) treats the entire repeated game as if it were T separate "subgames" (those occurring in rounds 1 modulo T, those occurring in rounds 2 modulo T, etc.), and it induces players to playing a grim-trigger strategy in each "subgame." Note, that agents need to know the calendar time perfectly, and that $T \to \infty$ when the players' discount factor converges to one.

[34] See Berger and Grüne (2014) who study observation of k actions, but restrict agents to play only image-scoring-like strategies.

defensive games is critical to the stability of cooperation when agents observe actions (and image scoring fails in offensive Prisoner's Dilemma games).

Structured Populations Some researchers have studied the scope of cooperation in the case where players do not have any information about their current partner but the matching of agents is not uniformly random. That is, the population is assumed to have some structure such that some agents are more likely to be matched to some partners than to other partners. A few papers (e.g., van Veelen, García, Rand, and Nowak, 2012; Alger and Weibull, 2013) show that it is possible to sustain cooperation with no information about the partner's behavior if matching is sufficiently assortative; i.e., cooperators are more likely to interact with other cooperators.[35] Our paper shows that letting players observe the partner's behavior in two interactions is sufficient to sustain cooperation without assuming assortativity.

Models without Calendar Time. The current paper differs from most of the literature on community enforcement by having a model without a global time zero. To the best of our knowledge, Rosenthal (1979) is the first paper to present the notion of a steady-state Nash equilibrium in environments in which each player observes the partner's last action, and apply it to the study of the Prisoner's Dilemma. Rosenthal focuses only on pure steady states (in which everyone uses the same pure strategy), and concludes that defection is the unique pure stationary Nash equilibrium action except in a few knife-edge cases. The methodology is further developed in Okuno-Fujiwara and Postlewaite (1995). Other papers following a related approach include Rubinstein and Wolinsky (1985), who study bargaining, Phelan and Skrzypacz (2006) who study repeated games with private monitoring, and Eliaz and Rubinstein (2014) who study boundedly rational agents. Our methodological contribution to the previous literature is that (1) we allow each agent to observe the behavior of the partner in several past interactions with other opponents, and (2) we combine the steady-state analysis with the presence of a few committed agents and present a novel notion of a perfect equilibrium to analyze this setup.

7.2 Empirical Predictions and Experimental Verification

In this section we discuss a few testable empirical predictions of our model, and comment on how to evaluate these predictions in lab experiments.

An experimental setup to evaluate our predictions would include a large group of subjects (say, at least 10) who play a large number of rounds (say, in expectation at least 50 rounds), and are rematched in each period to play a Prisoner's Dilemma game with a new partner. The experiment would include various treatments that differ in terms of (1) the parameters of the underlying game, e.g., whether the game is offensive/defensive and mild/acute, and (2) the information each agent observes about her partner: in particular, the number of past interactions that each agent observes, and what she observes in each interaction (e.g., actions, conflicts, or action profiles.)

Our theoretical predictions deal with a "pure" setup in which all agents maximize their material payoffs except for a vanishingly small number of committed agents. An experimental setup (and, arguably, real-life interactions) differs in at least two key respects: (1) agents, while caring about their material payoffs, may consider other non-material aspects, such as fairness and reciprocity, and (2) agents occasionally make mistakes and the frequency of these mistakes, while relatively low, is not negligible. In what follows, we describe our

[35] See also the following papers that study the stability of cooperation in other kinds of structured populations: Herold (2012) who studies a "haystack" model in which individuals interact within separate groups, Fujiwara-Greve and Okuno-Fujiwara (2009) who study a "voluntarily separable" repeated Prisoner's Dilemma, and Cooper and Wallace (2004) who study group selection.

key predictions in the "pure" setup, interpret its implications in a "noisy" experimental setup, and describe the relevant existing data.

Our first prediction (Theorems 1 and 2) deals with observation of the partner's actions, and it states that cooperation can be sustained only in defensive games. In an experimental setup we interpret this to imply that, *ceteris paribus*, the frequency of cooperation will be higher in a defensive game than in an offensive game. Engelmann and Fischbacher (2009), Molleman, van den Broek, and Egas (2013), and Swakman, Molleman, Ule, and Egas (2016) study the rate of cooperation in the borderline case of $g = l$ and in the closely related donor-recipient game, in which at each interaction only one of the players (the donor) chooses whether to give up g of her own payoff to yield a gain of $1 + g$ for the recipient. The typical findings in these experiments are that observation of 3–6 past actions induces a relatively high level of cooperation (50%–75%, where higher rates of cooperation are typically associated with environments in which more past actions are observed, and environments in which subjects can also observe second-order information about the behavior of the partner's past opponent). We are aware of only a single experiment that studies a setup in which $g \neq l$. Gong and Yang (2014) study the case of $g = 0.8 > l = 0.4$, and present results that seem to be consistent with our prediction. They observe an average rate of cooperation of only 30%–50%, even though in their setup players observe 10 past actions of the partner, and, in addition, are also able to observe the signal observed by the partner in each of these past interactions ("second-order information," which facilitates cooperation relative to the model analyzed in this paper).

Our second prediction (Theorems 3 and 4) deals with observation of either past conflicts or past action profiles, and it states that cooperation can be sustained only in mild games. In an experimental setup it implies that, *ceteris paribus*, the frequency of cooperation will be higher in mild games than in acute games. We are unaware of any existing experimental data with observation of either action profiles or conflicts.

It is interesting to compare our first two predictions to the comparative statics recently developed for repeated Prisoner's Dilemma games played by the same pair of players. Blonski, Ockenfels, and Spagnolo (2011), Dal Bó and Fréchette (2011), and Breitmoser (2015) present theoretical arguments and experimental data to suggest that when a pair of players repeatedly play the Prisoner's Dilemma, then the lower the values of g and l are, the easier it is to sustain cooperation.[36] However, our prediction is that when agents are randomly matched in each round, then the lower the value of g, and the *higher* the value of l, the easier it is to sustain cooperation.

Our final prediction is that when communities succeed in sustaining cooperation, it will be supported by the following behavior: most subjects defect (resp., cooperate, mix) when observing 2+ (resp., 0, 1) defections/conflicts. In an experimental setup we interpret this to predict that the probability that an agent defects increases with the number of times she observes the partner to be involved in defections/conflict. In particular, we predict a substantial increase in a subject's propensity to defect when moving from zero to two observations of defection. The findings of Engelmann and Fischbacher (2009), Molleman, van den Broek, and Egas (2013), Gong and Yang (2014), and Swakman, Molleman, Ule, and Egas (2016) suggest that subjects are indeed more likely to defect when they observe the partner to defect more often in the past.

7.3 Robustness of Results

In this section we discuss the robustness of our results with respect to various factors.

[36] Specifically, the above papers show that cooperation is more likely to be sustained in the infinitely repeated Prisoner's Dilemma if the discount factor of the players is above $\frac{g+l}{g+l+1}$. Note that this minimal threshold for cooperation is increasing in both parameters. Embrey, Frechette, and Yuksel (2015) present similar comparative statics evidence for the finitely repeated Prisoner's Dilemma.

Joint Deviations and Evolutionary Stability Our main solution concept (namely, perfect equilibrium) considers only deviations by a single agent (who has mass zero in the infinite population). A stronger solution concept is the notion of evolutionarily stable strategy (Maynard Smith and Price, 1973) that requires stability also against a group of agents with a positive small mass who jointly deviate. This stronger notion also implies asymptotic stability in many monotonic dynamics (see, e.g., Sandholm, 2010); that is, if due to a small perturbation the population state slightly moves away from an evolutionarily stable state, then plausible dynamics, in which more successful strategies become more frequent, will take the population back to that state.

In Appendix B we adapt the notion of an evolutionarily stable strategy to the setup of an environment with observation of the partner's past behavior, and we show that the perfect equilibria that support cooperation in our main results satisfy the refinement of evolutionary stability.

Cheap Talk and Equilibrium Selection Appendix C shows that both cooperation and defection are evolutionarily stable in defensive games with observation of actions. In Appendix C we study the influence of the introduction of pre-play "cheap-talk" communication on this stability result. Specifically, we assume that there are slightly costly signals that, due to their positive cost, are not used unless they yield a benefit. In this setup one can obtain the following results.

1. Offensive games: No stable state exists. Both defection and cooperation are only "quasi-stable"; the population state occasionally changes between theses two states, based on the occurrence of rare random experimentations. The argument is adapted from Wiseman and Yilankaya (2001).

2. Defensive games (and $k \geq 2$): The introduction of cheap talk destabilizes all non-efficient equilibria, leaving cooperation as the unique stable outcome. The argument is adapted from Robson (1990).

General Noise Structures In the model described above we deal with perturbed environments that include a single kind of noise, namely, committed agents who follow commitment strategies. It is possible to extend our results to include additional sources of noise: specifically, observation noise and/or trembles. We redefine a perturbed environment as a tuple $E_{\epsilon,\delta} = \left((G,k), \left(S^C, \lambda\right), \alpha, \epsilon, \delta\right)$, where $(G,k), \left(S^C, \lambda\right), \epsilon$ are defined as in the main model, $0 < \delta << 1$ is the probability of error in each observed action of a player, and $\alpha \in \Delta(A)$ is a totally mixed distribution according to which the observed error is sampled from in the event of an observation error. Alternatively, these errors can also be interpreted as actions played by mistake by the partner due to trembling hands. The notion of a steady state (and, in particular, consistent behavior) can be adapted to the setup with observation errors in a straightforward way. Indeed, one can show that *all* of our results can be adapted to this setup in a relatively straightforward way. In particular, our results hold also in environments in which most of the noise is due to observation errors, provided that there is a small positive share of committed agents (possibly much smaller than the probability of an observation error).[37]

Hybrid Model: Finitely Lived Agents With no Global Time Zero A previous version of this manuscript dealt with finitely lived agents in a model with no global time zero (omitted for brevity in the current version). Specifically, the alternative model combines the realistic assumptions of the conventional repeated game model

[37] Formally, one needs to redefine a perfect equilibrium as the limit of Nash equilibria in a converging sequence of perturbed environments $\left((G,k), \left(S^C, \lambda\right), \alpha, \epsilon_n, \delta_n\right)$ where $\epsilon_n, \delta_n \to 0$. Next, we say that action $a \in A$ is a strictly perfect equilibrium in this extended setup if for any converging sequence of perturbed environments $\left((G,k), \left(S^C, \lambda\right), \alpha, \epsilon_n, \delta_n\right)$ satisfying $\epsilon_n, \delta_n \to 0$ and $\frac{\epsilon_n}{\delta_n} \to$ constant (which is allowed to be 0 or ∞), there exists a converging sequence of Nash equilibria $\left(S_n^N, \sigma_n, \theta_n\right) \to (S^*, \sigma^*, \theta^* \equiv a)$ such that their outcomes converge to an outcome in which all normal agents play action a with probability one.

of Section 6 (agents are finitely lived, allowed to choose non-stationary strategies, and observe the partner's recent actions), while relaxing the assumption that the entire community started interacting in a specific round (global time zero). The alternative model assumes that in each round a small share $1-\delta$ of the agents are chosen at random to stop interacting (retire) and are replaced with new agents, who begin interacting with an empty history of past actions. When two agents interact, each of them observes the last k actions played by her partner (or the entire partner's history, if the partner is "young" and has interacted in fewer than k rounds).

One can show that our results can be adapted to this setup in a similar manner to their adaptation to the conventional setup of repeated games. Namely, most of our results hold in this alternative setup except for the three points discussed in Section 6.3.

7.4 Conclusion and Directions for Future Research

In many situations people engage in short-term interactions where they are tempted to behave opportunistically but there is a possibility that future partners will obtain some information about their behavior today. We propose a new modeling approach based on the premises that (1) an equilibrium has to be robust to the presence of a few committed agents, and (2) the community has been interacting from time immemorial (though this latter assumption is relaxed in Section 6).

We develop a novel methodology that allows for a tractable analysis of these seemingly complicated environments. We apply this methodology to the study of Prisoner's Dilemma games (and coordination games), and we obtain sharp testable predictions for the equilibrium outcomes, and the exact conditions under which cooperation can be sustained as an equilibrium outcome. Finally, we show that whenever cooperation is sustainable, there is a unique (and novel) way to support it that has a few appealing properties: (1) agents behave in an intuitive and simple way, and (2) the equilibrium is robust, e.g., to deviations by a group of agents, or to the presence of any kind of committed agents.

We believe that our novel modeling approach will be helpful in understanding various interactions in future research. In particular, we plan to extend the methodology to asymmetric games. Another direction for future research is to adapt the model to better fit online interactions, and to deal with non-verifiable public reports similar to the online feedback mechanisms in web-sites such as eBay. Finally, readers may be interested in our companion paper (Heller and Mohlin, 2016a), in which we study a related setup in which agents are allowed to exert effort in deception by influencing the signal observed by the opponent.

All the standard equilibrium refinements of extensive-form games (e.g., "trembling-hand" perfection, sequential equilibrium, and subgame perfection) assume (implicitly or explicitly) that any observation of behavior that is inconsistent with the equilibrium strategies is due to a "tremble." Consequently an agent believes with probability one that the partner will follow the equilibrium in the future regardless of how many times the partner has deviated in the past. We think that this assumption may be problematic in various applications. In this paper we present an alternative approach, according to which an agent who has deviated from the equilibrium behavior in the past is more likely to be a committed agent who may deviate from the equilibrium behavior again in the future. We think that this approach may be helpful in future research to develop a new equilibrium refinement that can be applied to various extensive-form games (possibly, this notion may be the related "normal-form" perfection).

References

Alger, I., and J. W. Weibull (2013): "Homo Moralis – Preference evolution under incomplete information and assortative matching," *Econometrica*, 81(6), 2269–2302.

Berger, U., and A. Grüne (2014): "Evolutionary stability of indirect reciprocity by image scoring," *mimeo*.

Bernstein, L. (1992): "Opting out of the legal system: Extralegal contractual relations in the diamond industry," *The Journal of Legal Studies*, 21(1), 115–157.

Bhaskar, V., G. J. Mailath, and S. Morris (2013): "A foundation for Markov equilibria in sequential games with finite social memory," *The Review of Economic Studies*, 80(3), 925–948.

Blonski, M., P. Ockenfels, and G. Spagnolo (2011): "Equilibrium selection in the repeated prisoner's dilemma: Axiomatic approach and experimental evidence," *American Economic Journal: Microeconomics*, 3(3), 164–192.

Breitmoser, Y. (2015): "Cooperation, but no reciprocity: Individual strategies in the repeated prisoner's dilemma," *American Economic Review*, 105(9), 2882–2910.

Celetani, M., D. Fudenberg, D. K. Levine, and W. Pesendorfer (1996): "Maintaining a reputation against a long-lived opponent," *Econometrica*, 64(3), 691–704.

Cooper, B., and C. Wallace (2004): "Group selection and the evolution of altruism," *Oxford Economic Papers*, 56(2), 307–330.

Cripps, M. W., G. J. Mailath, and L. Samuelson (2004): "Imperfect monitoring and impermanent reputations," *Econometrica*, 72(2), 407–432.

Dal Bó, P., and G. R. Fréchette (2011): "The evolution of cooperation in infinitely repeated games: Experimental evidence," *The American Economic Review*, 101(1), 411–429.

Deb, J. (2012): "Cooperation and community responsibility: A folk theorem for repeated matching games with names," *Available at SSRN 1213102*.

Deb, J., and J. González-Díaz (2014): "Community enforcement beyond the prisoner's dilemma," *mimeo*.

Dixit, A. (2003): "On modes of economic governance," *Econometrica*, 71(2), 449–481.

Duffy, J., and J. Ochs (2009): "Cooperative behavior and the frequency of social interaction," *Games and Economic Behavior*, 66(2), 785–812.

Eliaz, K., and A. Rubinstein (2014): "A model of boundedly rational 'neuro' agents," *Economic Theory*, 57(3), 515–528.

Ellison, G. (1994): "Cooperation in the prisoner's dilemma with anonymous random matching," *The Review of Economic Studies*, 61(3), 567–588.

Ely, J., D. Fudenberg, and D. K. Levine (2008): "When is reputation bad?," *Games and Economic Behavior*, 63(2), 498–526.

Embrey, M., G. R. Frechette, and S. Yuksel (2015): "Cooperation in the finitely repeated prisoner's dilemma," *mimeo*.

Engelmann, D., and U. Fischbacher (2009): "Indirect reciprocity and strategic reputation building in an experimental helping game," *Games and Economic Behavior*, 67(2), 399–407.

Fudenberg, D., and D. K. Levine (1989): "Reputation and equilibrium selection in games with a patient player," *Econometrica*, 57(4), 759–778.

Fujiwara-Greve, T., and M. Okuno-Fujiwara (2009): "Voluntarily separable repeated prisoner's dilemma," *The Review of Economic Studies*, 76(3), 993–1021.

Gong, B., and C.-L. Yang (2014): "Reputation and cooperation: An experiment on prisoner's dilemma with second-order information," *mimeo*.

Greif, A. (1993): "Contract enforceability and economic institutions in early trade: The Maghribi traders' coalition," *The American Economic Review*, pp. 525–548.

Heller, Y. (2015): "Three steps ahead," *Theoretical Economics*, 10, 203–241.

——— (2017): "Instability of Belief-Free Equilibria," *Journal of Economic Theory*, 168, 261–286, Mimeo.

Heller, Y., and E. Mohlin (2016a): "Coevolution of deception and preferences: Darwin and Nash meet Machiavelli," *mimeo*, Mimeo.

——— (2016b): "Unique Stationary Behaivor," *mimeo*, Mimeo.

Herold, F. (2012): "Carrot or stick? The evolution of reciprocal preferences in a Haystack model," *American Economic Review*, 102(2), 914–940.

Herold, F., and C. Kuzmics (2009): "Evolutionary stability of discrimination under observability," *Games and Economic Behavior*, 67(2), 542–551.

Jøsang, A., R. Ismail, and C. Boyd (2007): "A survey of trust and reputation systems for online service provision," *Decision Support Systems*, 43(2), 618–644.

Kandori, M. (1992): "Social norms and community enforcement," *The Review of Economic Studies*, 59(1), 63–80.

Kim, Y.-G., and J. Sobel (1995): "An evolutionary approach to pre-play communication," *Econometrica*, 63(5), 1181–1193.

Kohlberg, E., and J.-F. Mertens (1986): "On the strategic stability of equilibria," *Econometrica*, 54(5), 1003–1037.

Kreps, D. M., P. Milgrom, J. Roberts, and R. Wilson (1982): "Rational cooperation in the finitely repeated prisoners' dilemma," *Journal of Economic Theory*, 27(2), 245–252.

Leimar, O., and P. Hammerstein (2001): "Evolution of cooperation through indirect reciprocity," *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 268(1468), 745–753.

Mailath, G. J., and L. Samuelson (2006): *Repeated games and reputations*, vol. 2. Oxford University Press.

Matsushima, H., T. Tanaka, and T. Toyama (2013): "Behavioral Approach to Repeated Games with Private Monitoring IRJE-F-879," University of Tokyo Faculty of Economics Discussion paper.

Maynard Smith, J. (1974): "The theory of games and the evolution of animal conflicts," *Journal of Theoretical Biology*, 47(1), 209–221.

Maynard Smith, J., and G. R. Price (1973): "The logic of animal conflict," *Nature*, 246, 15.

Milgrom, P., D. C. North, and B. R. Weingast (1990): "The role of institutions in the revival of trade: The law merchant, private judges, and the Champagne fairs," *Economics and Politics*, 2(1), 1–23.

Molleman, L., E. van den Broek, and M. Egas (2013): "Personal experience and reputation interact in human decisions to help reciprocally," *Proceedings of the Royal Society of London B: Biological Sciences*, 280(1757), 20123044.

Nowak, M. A., and K. Sigmund (1998): "Evolution of indirect reciprocity by image scoring," *Nature*, 393(6685), 573–577.

Ohtsuki, H., and Y. Iwasa (2006): "The leading eight: Social norms that can maintain cooperation by indirect reciprocity," *Journal of Theoretical Biology*, 239(4), 435–444.

Okada, A. (1981): "On stability of perfect equilibrium points," *International Journal of Game Theory*, 10(2), 67–73.

Okuno-Fujiwara, M., and A. Postlewaite (1995): "Social norms and random matching games," *Games and Economic Behavior*, 9(1), 79–109.

Osborne, M. J., and A. Rubinstein (1994): *Course in Game Theory*. MIT Press.

Panchanathan, K., and R. Boyd (2003): "A tale of two defectors: The importance of standing for evolution of indirect reciprocity," *Journal of Theoretical Biology*, 224(1), 115–126.

Phelan, C., and A. Skrzypacz (2006): "Private monitoring with infinite histories," Discussion paper, Federal Reserve Bank of Minneapolis.

Resnick, P., and R. Zeckhauser (2002): "Trust among strangers in Internet transactions: Empirical analysis of eBay's reputation system," *The Economics of the Internet and E-commerce*, 11(2), 23–25.

Robson, A. J. (1990): "Efficiency in evolutionary games: Darwin, Nash, and the secret handshake," *Journal of Theoretical Biology*, 144(3), 379–396.

Rosenthal, R. W. (1979): "Sequences of games with varying opponents," *Econometrica*, 47(6), 1353–1366.

Rubinstein, A., and A. Wolinsky (1985): "Equilibrium in a market with sequential bargaining," *Econometrica*, 53(5), 1133–1150.

Sakovics, J., and J. Steiner (2012): "Who matters in coordination problems?," *The American Economic Review*, 102(7), 3439–3461.

Sandholm, W. H. (2010): "Local stability under evolutionary game dynamics," *Theoretical Economics*, 5(1), 27–50.

Schlag, K. H. (1993): "Cheap talk and evolutionary dynamics," Bonn Department of Economics Discussion Paper B-242.

Selten, R. (1975): "Reexamination of the perfectness concept for equilibrium points in extensive games," *International Journal of Game Theory*, 4(1), 25–55.

——— (1980): "A note on evolutionarily stable strategies in asymmetric animal conflicts," *Journal of Theoretical Biology*, 84(1), 93–101.

——— (1983): "Evolutionary stability in extensive two-person games," *Mathematical Social Sciences*, 5(3), 269–363.

Sugden, R. (1986): *The Economics of Rights, Co-operation and Welfare*. Blackwell Oxford.

Swakman, V., L. Molleman, A. Ule, and M. Egas (2016): "Reputation-based cooperation: Empirical evidence for behavioral strategies," *Evolution and Human Behavior*, 37(3), 230–235.

Takahashi, S. (2010): "Community enforcement when players observe partners' past play," *Journal of Economic Theory*, 145(1), 42–62.

Thomas, B. (1985): "On evolutionarily stable sets," *Journal of Mathematical Biology*, 22(1), 105–115.

van Veelen, M., J. García, D. G. Rand, and M. A. Nowak (2012): "Direct reciprocity in structured populations," *Proceedings of the National Academy of Sciences*, 109(25), 9929–9934.

Weibull, J. W. (1995): *Evolutionary Game Theory*. MIT Press.

Wiseman, T., and O. Yilankaya (2001): "Cooperation, secret handshakes, and imitation in the prisoners' dilemma," *Games and Economic Behavior*, 37(1), 216–242.

A Example: Perfect Equilibrium with Partial Cooperation (Online Publication)

The following example demonstrates the existence of a *non-regular* perfect equilibrium of an offensive Prisoner's Dilemma, in which players cooperate with positive probability.

Example 6 (Non-regular Perfect Equilibrium with Partial Cooperation)**.** Consider the environment $(G_O, 1)$ where G_O is an offensive Prisoner's Dilemma game with $g = 2.3$, $l = 1.7$ (see Table 1), and each agent observes a single action sampled from the partner's behavior. Let s^* be the strategy that defects with probability 10% after observing cooperation (i.e., $m = 0$) and defects with probability 81.7% (numerical values in this example are rounded to 0.1%) after observing a defection (i.e., $m = 1$). Let q^* denote the average probability of defection in a homogeneous population of agents who follow strategy s^*. The value of q^* is calculated as follows:

$$q^* = (1 - q^*) \cdot 10\% + q^* \cdot 81.7\% \;\Rightarrow\; q^* = 35.3\%. \tag{13}$$

Eq. (13) holds because an agent defects in either of the following exhaustive cases: (1) she observes cooperation (which happens with a probability of $1 - q^*$) and then she defects with probability 10%, or (2) she observes defection (which happens with a probability of q^*) and then she defects with probability 81.7%. This implies that the unique consistent signal θ^* of a homogeneous population in which all agents follow s^* satisfies $\theta^*(1) = 35.3\%$ (i.e., agents defect in 35.3% of the observed interactions).
Next, observe that an agent who follows strategy s^* defects with probability

$$p(q) = q \cdot 81.7\% + (1 - q) \cdot 10\%$$

when being matched with a partner who defects with an average probability of q. This implies that the payoff of a deviator (Alice) who defects with an average probability of q is

$$\pi_q\left(\left(\{s^*\}, 1_{s^*}, \theta^*\right)\right) = q \cdot (1 - p(q)) \cdot (1 + g) + (1 - q) \cdot p(q) \cdot (-l) + (1 - q) \cdot (1 - p(q)) \cdot 1.$$

This is because with a probability of $q \cdot (1 - p(q))$ only Alice defects, with a probability of $(1 - q) \cdot p(q)$ only Alice cooperates, and with a probability of $(1 - q) \cdot (1 - p(q))$ both players cooperate. By calculating the FOC one can show that $q = q^* = 35.3\%$ is the probability of defection that uniquely maximizes the payoff of a deviator. This implies that $(\{s^*\}, 1_{s^*}, 35.3\%)$ is a Nash equilibrium of the (non-regular) perturbed environments $(G, k, \{s^*\}, 1_{s^*}, \epsilon)$ for any $\epsilon \in (0, q)$, which implies that $(\{s^*\}, 1_{s^*}, \theta^*)$ is a (non-regular) perfect equilibrium.
The above perfect equilibrium relies on a very particular set of commitment strategies in which all committed agents happen to play the same strategy as the normal agents. This cannot hold in a regular set of commitment strategies, in which different commitment strategies defect with different average probabilities. Given this regularity, it must be the case that the conditional probability that the partner is going to defect is higher after he observes a defection ($m = 1$) than after he observes a cooperation ($m = 0$). This implies that a deviator (Alice) who defects with a probability of 35.3% regardless of the signal will strictly outperform the incumbents. This is because the incumbents behave the same against Alice (as she has the same average probability of defection as the incumbents), while Alice defects with higher probability against partners who are more likely to cooperate (i.e., after she observes $m = 0$), which implies that due to the offensiveness of the game (i.e., $g > l$), Alice achieves a strictly higher payoff than the incumbents.

B Evolutionary Stability (Online Publication)

In the main text we have dealt with the notion of perfect equilibrium, which requires that no agent be able to achieve a better payoff than the incumbents by unilateral deviation. In this appendix we refine the solution concept to require stability also against small groups of agents (with a positive small mass) who deviate together. It turns out that all of our results also work under this refinement.

B.1 Definitions

In a seminal paper, Maynard Smith and Price (1973) define a symmetric Nash equilibrium strategy α^* to be evolutionarily stable if the incumbents achieve a strictly higher payoff when being matched with any other best-reply strategy β (i.e., $\pi(\beta,\alpha^*)=\pi(\alpha^*,\alpha^*) \Rightarrow \pi(\alpha^*,\beta)>\pi(\beta,\beta)$. The motivation is that if β is a best reply to α^*, then a single deviator who plays β will be as successful as the incumbents. This may induce a few other agents to mimic her behavior, until a small positive mass of agents follow β. The above inequality implies that at this stage the followers of β will be strictly outperformed, and thus will disappear from the population.

Our setup with environments is similar to the standard setup of a repeated game in that it rarely admits evolutionarily stable strategies. Typically, not all the actions will be played by normal agents in equilibrium, and as a result some signals will never be observed. Deviators who differ in their behavior only after such zero probability signals will get the same payoff as the incumbents both against the incumbents and against other deviators. This violates the above inequality.

Following Selten's (1983) notion of "limit ESS," we solve this issue by requiring evolutionary stability in a converging sequence of perturbed environments, in which all signals are observed on the equilibrium path, instead of simply requiring evolutionary stability in the unperturbed environment.

This is formalized as follows. Given a steady state (S,σ,θ) in a perturbed environment $\left((G,k),\left(S^C,\lambda\right),\epsilon\right)$, we define $\pi_{\hat{s}}(\hat{s})$ as the (long-run average) payoff of strategy $\hat{s}$ against itself, and $\pi_{(S,\sigma)}(\hat{s})$ as the mean (long-run average) payoff of the incumbents against strategy $\hat{s}$. Specifically, if $\hat{s}\in S\cup S^C$, then

$$\pi_{\hat{s}}(\hat{s}|S,\sigma,\theta)=\sum_{(a,a')\in A^2}\hat{\theta}_{\hat{s}}(\hat{s})(a)\cdot\hat{\theta}_{\hat{s}}(\hat{s})(a')\cdot\pi(a,a'),$$

$$\pi_{(S,\sigma)}(\hat{s}|S,\sigma,\theta)=\sum_{s\in S\cup S^C}\sum_{(a,a')\in A^2}((1-\epsilon)\cdot\sigma(s)+\epsilon\cdot\lambda(s))\cdot\theta_s(\hat{s})(a)\cdot\hat{\theta}_{\hat{s}}(s)(a')\cdot\pi(a,a'),$$

and if $\hat{s}\notin S\cup S^C$, then we define $\pi_{\hat{s}}(\hat{s})$ and $\pi_{(S,\sigma)}(\hat{s})$ as the respective payoffs in the post-deviation steady state $\left(S\cup\{\hat{s}\},\hat{\sigma},\hat{\theta}\right)$:

$$\pi_{\hat{s}}(\hat{s}|S,\sigma,\theta)=\sum_{(a,a')\in A^2}\hat{\theta}_{\hat{s}}(\hat{s})(a)\cdot\hat{\theta}_{\hat{s}}(\hat{s})(a')\cdot\pi(a,a'),$$

$$\pi_{(S,\sigma)}(\hat{s}|S,\sigma,\theta)=\sum_{s\in S\cup S^C}\sum_{(a,a')\in A^2}((1-\epsilon)\cdot\hat{\sigma}(s)+\epsilon\cdot\lambda(s))\cdot\hat{\theta}_s(\hat{s})(a)\cdot\hat{\theta}_{\hat{s}}(s)(a')\cdot\pi(a,a').$$

Definition 15. A steady state (S^*,σ^*,θ^*) of a perturbed environment $\left((G,k),\left(S^C,\lambda\right),\epsilon\right)$ is *evolutionarily stable* if (1) (S^*,σ^*,θ^*) is a Nash equilibrium, and (2) for any best-reply strategy $\hat{s}$ (i.e., $\pi_{\hat{s}}(S^*,\sigma^*,\theta^*)=\pi(S^*,\sigma^*,\theta^*)$), such that $\sigma^*(\hat{s})<1$ (i.e., $\hat{s}$ is not the only normal strategy) the following inequality holds: $\pi_{(S,\sigma)}(\hat{s}|S,\sigma,\theta)>\pi_{\hat{s}}(\hat{s}|S,\sigma,\theta)$.

Definition 16. A steady state (S^*,σ^*,θ^*) of the environment (G,k) is *a perfect evolutionarily stable state* if there exist a distribution of commitments $\left(S^C,\lambda\right)$ and converging sequences $\left(S_n^N,\sigma_n,\theta_n\right)_n\to_{n\to\infty}(S^*,\sigma^*,\theta^*)$

and $(\epsilon_n > 0)_n \to_{n\to\infty} 0$, such that for each n, the state $\left(S_n^N, \sigma_n, \theta_n\right)$ is an evolutionarily stable state in the perturbed environment $\left((G,k), \left(S^C, \lambda\right), \epsilon_n\right)$. If the outcome assigns probability one to one of the actions, i.e., $\theta^* \equiv a$, then we say that this action is a perfect evolutionarily stable outcome.

Finally, we define a strictly perfect evolutionarily stable outcome as a pure action that is an outcome of a perfect evolutionarily stable state for any distribution of commitments (similar to the notion of strict limit ESS in Heller, 2015).

Definition 17. Action $a^* \in A$ is a *strictly perfect evolutionarily stable outcome* in the environment $E = ((A,\pi),k)$ if, for any distribution of commitment strategies $\left(S^C, \lambda\right)$, there exist a steady state $(S^*, \sigma^*, \theta^* \equiv a^*)$ and converging sequences $\left(S_n^N, \sigma_n, \theta_n\right)_n \to_{n\to\infty} (S^*, \sigma^*, \theta^*)$ and $(\epsilon_n > 0)_n \to_{n\to\infty} 0$, such that for each n, the state $\left(S_n^N, \sigma_n, \theta_n\right)$ is an evolutionarily stable state in the perturbed environment $\left((G,k), \left(S^C, \lambda\right), \epsilon_n\right)$.

B.2 Adaptation of Results

All of our results hold with respect to the refinement of evolutionary stability. In particular, the fact that always defecting is a strict equilibrium (i.e., the unique best reply to itself) in any slightly perturbed environment implies that defection is a strictly perfect evolutionarily stable outcome.

Similarly, one can adapt the results about sustaining cooperation as an equilibrium action (Theorems 2–5). Specifically, minor modifications to the proofs can show that cooperation is a strictly perfect evolutionarily stable outcome in defensive games with observation actions and in mild games with observation of conflicts (when $k \geq 2$), and that cooperation is a perfect evolutionarily stable outcome in mild games with observation of action profiles, and in any game with observation of actions against defectors.

A sketch of the argument why the results apply also to the refinement of evolutionary stability is as follows. There are two kinds of steady states that sustain cooperation in the proofs in this paper:

1. Steady state $\psi'_n = (s^{q_n}, 1_{s^{q_n}}, \theta_n)$ that has a single normal strategy in its support. The arguments in the proofs show that each such strategy is the unique best reply to itself in the n^{th} perturbed environment (i.e., $\pi_{s'}(\psi'_n) < \pi(\psi'_n)$ for each $s' \neq s^{q_n}$), which shows that ψ'_n is an evolutionarily stable state in the n^{th} perturbed environment.

2. Steady state $\psi_n = \left(\left\{s^1, s^2\right\}, (q_n, 1-q_n), \theta_n\right)$ that consists of two normal strategies in its support. The arguments in the proofs show that these two strategies are the only best replies to this steady state (i.e., $\pi_{s'}(\psi_n) < \pi(\psi_n\theta_n)$ for each $s' \notin \left\{s^1, s^2\right\}$). Moreover, the arguments in the proof (see, in particular, Remark 12 at the end of the proof of Theorem 2) imply that each of these two normal strategies obtains a relatively low payoff when being matched against itself, i.e.,: $\pi\left(s^1|\psi_n\right) > \pi_{s^1}\left(s^1|S,\sigma,\theta\right)$ and $\pi\left(s^2|\psi_n\right) > \pi_{s^2}\left(s^2|\psi_n\right)$, which implies that ψ_n is evolutionarily stable.

C Cheap Talk and Equilibrium Selection (Online Publication)

Appendix B shows that both perfect equilibrium outcomes, namely, cooperation and defection, satisfy the refinement of evolutionary stability. In this section we discuss how the stability analysis changes if one introduces pre-play "cheap-talk" communication in our setup.

For concreteness, we focus on observation of actions. As in the standard setup of normal-form games (without observation of past actions), the introduction of cheap talk induces different equilibrium selection results, depending on whether or not deviators have unused signals to use as secret handshakes (see, e.g.,

Robson, 1990; Schlag, 1993; Kim and Sobel, 1995). If one assumes that the set of cheap-talk signals is finite, and all signals are costless, then cheap talk has little effect on the set of perfect equilibrium outcomes (as any perfect equilibrium of the game without cheap talk can be implemented as an equilibrium with cheap talk in which the incumbents send all signals with positive probability).

In what follows we focus on a different case, in which there are slightly costly signals that, due to their positive cost, are not used unless they yield a benefit. In this setup our results should be adapted as follows.

1. Offensive games: No stable state exists. Both defection and cooperation are only "quasi-stable"; the population state occasionally changes between theses two states, based on the occurrence of rare random experimentations. The argument is adapted from Wiseman and Yilankaya (2001).

2. Defensive games (and $k \geq 2$): The introduction of cheap talk destabilizes all non-efficient equilibria, leaving cooperation as the unique stable outcome. The argument is adapted from Robson (1990).

In what follows we only briefly sketch the arguments for these results, since a formal presentation would be very lengthy, and the contribution is somewhat limited given that similar arguments have already been presented in the literature.

Following Wiseman and Yilankaya (2001), we modify the environment by endowing agents with the ability to send a slightly costly signal ϕ (called the *secret handshake*). An agent has to pay a small cost c either to send ϕ to her partner or to observe whether the partner has sent ϕ to her. In addition, we still assume that each agent observes $k \geq 2$ past actions of the partner. Let ξ be the initial small frequency of a group of experimenting agents (called *mutants*) who deviate jointly. We assume that $O(\epsilon) \cdot O(\xi) < c < O(\xi)$, i.e., that the small cost of the secret handshake is smaller than the initial share of mutants, but larger than the product of the two small shares of the mutants ($O(\xi)$) and the committed agents ($O(\epsilon)$). To simplify the analysis we also assume that the committed agents do not use the secret handshake

Consider a population that starts at the defection equilibrium, in which all normal agents defect regardless of the observed actions and do not use signal ϕ. Consider a small group of ξ mutants ("cooperative handshakers") who send the signal ϕ, and cooperate iff the partner has sent ϕ as well. These mutants outperform the incumbents: they achieve ξ additional points by cooperating among themselves, which outweighs the cost of $2 \cdot c$ for using the secret handshake. Thus, assuming a payoff-monotonic selection dynamics, the mutants take over the population and destabilize the defective equilibrium. If the underlying game is offensive, then there is no other candidate to be a stable population state. Thus, cooperation can be sustained only until new mutants arrive ("defective handshakers") who use the secret handshake and always defect. These mutants outperform the cooperative handshakers, and would take over the population. Finally, a third group of mutants who always defect without using the secret handshake can take the population back to the starting point.

If the underlying game is defensive, then there is a sequence of mutants who can take the population into the cooperative equilibrium characterized in the main text. Specifically, the second group of mutants (the ones after the cooperative handshakers) include agents who send only ϕ, but instead of incurring the small cost c of observing the partner's secret handshake, they base their behavior on the partner's observed actions, namely, they play some combination of the strategies s^1 and s^2. This second group of mutants would take over the population because the cost they save by not checking the secret handshake outweighs the small loss of $O(\epsilon)$ incurred from not defecting against committed partners. Finally, a third group of mutants who do not send the secret handshake, and follow strategies s^1 and s^2, can take over the population (by saving the cost of sending ϕ), and induce the perfect cooperative equilibrium of the main text. This equilibrium remains stable also with the option of using the secret handshake because (1) mutants who defect when observing $m = 0$ are outperformed

due to similar arguments to those in the main model, and (2) mutants who send the secret handshake, and always cooperate when observing ϕ (also when $m > 2$), are outperformed, as the cost of the secret handshake c outweighs the gain of $O(\xi) \cdot O(\epsilon)$.

D Proofs (Online Publication)

D.1 Proof of Proposition 1 (Implementation of Perfect Equilibria)

If α is a totally mixed strategy, then it is immediate that the state $(\{\alpha\}, \nu_\alpha)$ is a Nash equilibrium of the perturbed environment$((G,k), \{\alpha\}, \epsilon)$ for any $\epsilon > 0$, which implies that the state $(\{\alpha\}, \nu_\alpha)$ is a perfect equilibrium. Assume now that α is not totally mixed. The fact that $\alpha \in \Delta(A)$ is a symmetric perfect equilibrium of the underlying game implies (see Selten, 1975, Theorem 7) that there is a sequence of totally mixed strategies $(\alpha_n) \to_{n\to\infty} \alpha$, such that α is a best reply to each α_n. The fact that α is a best reply both to itself and to α_1 (the first element in the sequence (α_n)) implies that the state $(\{\alpha\}, \nu_\alpha)$ is a Nash equilibrium of the regular perturbed environment $((G,k), (\{\alpha_1, \alpha\}, (0.5, 0.5)), \epsilon)$ for any $\epsilon > 0$, which implies that $(\{\alpha\}, \nu_\alpha)$ is a regular perfect equilibrium.

D.2 Proof of Proposition 2 (Mixed Equilibrium in Coordination Game)

Assume to the contrary that $(\{\alpha\}, \nu_\alpha)$ is a regular perfect equilibrium in the environment $(G, k \geq 1)$. This implies that $(\{\alpha\}, \nu_\alpha)$ is a Nash equilibrium of some regular perturbed environment $\left((G,k), \left(S^C, \lambda\right), \epsilon > 0\right)$. The regularity of $\left(S^C, \lambda\right)$ implies that there is $s \in S^C$ such that $s_{\nu_\alpha} \neq \alpha$. Assume w.l.o.g. that $s_{\nu_\alpha}(a) > \alpha(a)$. This inequality implies that when an agent observes a signal $m_{\vec{a}} = (a, ..., a)$ (i.e., the partner played the action a in all k observed interactions), then there is a posterior probability strictly larger than $\alpha(a)$ that the partner is going to play a. This implies that playing a when observing signal $m_{\vec{a}}$ induces a strictly larger payoff than playing α, which contradicts $(\{\alpha\}, \nu_\alpha)$ being a Nash equilibrium in the regular perturbed environment.

D.3 Proof of Proposition 3 (Strictly Perfect Outcomes in Coordination Games)

We identify each signal m with the number of times that action b has been played in the sample of k observations.

Case I: Suppose that $\pi(a,a) < \pi(b,b)$. We want to show that a is not a strictly perfect equilibrium action. Assume to the contrary that a is a strictly perfect equilibrium action. Let s^α be the strategy such that $s^\alpha_k(a) = \alpha$ (and $s^\alpha_k(b) = 1 - \alpha$) for all k. Pick $\alpha > 0$ sufficiently small such that b is the unique best reply against s^α. Consider a perturbed environment $((G,k), \{s^\alpha\}, \epsilon > 0)$. The assumption that a is strictly perfect implies that there is a steady state $(S^*, \sigma^*, \theta^* \equiv \nu_a)$, a converging sequence of steady states $\left(S^N_n, \sigma_n, \theta_n\right) \to (S^*, \sigma^*, \theta^*)$, and a converging sequence of perturbed environments $((G,k), (\{s^\alpha\}, 1_{s^\alpha}), \epsilon_n)$, such that each $\left(S^N_n, \sigma_n, \theta_n\right)$ is a Nash equilibrium of $((G,k), \{s^\alpha\}, \epsilon_n)$. Fix a sufficiently small ϵ_n (sufficiently large n).

Assume first that $(s_n)_k(b) = 1$ for each $s_n \in S^N_n$ (i.e., all normal agents play b with probability one if they observe only b's). This implies that a deviating agent (Alice) who always plays b outperforms the incumbents: Alice will get a high payoff very close to $\pi(b,b)$ (because both she and all of her normal partners play b), while the incumbents achieve a lower average payoff of about $\pi(a,a)$ (because $\theta_n \to_{n\to\infty} \theta^* \equiv a$). This contradicts that $\left(S^N_n, \sigma_n, \theta_n\right)$ is a Nash equilibrium of $((G,k), \{s^\alpha\}, \epsilon_n)$.

Next assume that there is a strategy $s_n \in S^N_n$ such that $(s_n)_k(b) < 1$. Note that when an agent observes signal $m = k$, it implies that with high probability the partner is following the commitment strategy s^α (because $\theta_n \to_{n\to\infty} \theta^* \equiv a$) , so that the unique "myopic" best reply (taking into account only the payoff in this

interaction, and not the fact that the action is observed by future partners) is action b. The fact that a normal agent who follows s_n plays a with positive probability when she observes signal $m = k$ implies that the direct loss of playing a when observing k must be compensated by the indirect gain accruing from interactions with future partners who observe the current interaction (otherwise $\left(S_n^N, \sigma_n, \theta_n\right)$ could not be a Nash equilibrium). This indirect future gain is independent of the current partner's behavior, while the direct loss from playing a is strictly larger when observing $m = k$ than when observing $m < k$. Hence, playing a is the unique best reply when an agent observes any signal $m < k$ (taking into account both the direct and the indirect impact of the played action on the payoff). In particular, all normal agents play a when observing $m = 1$. This implies that the indirect loss of playing b when observing $m = k$ is very small $\left(O\left(\epsilon_n^k\right)\right)$ because the probability of observing the signal $m = k$ is small $\left(O\left(\epsilon_n\right)\right)$, and hence it is very unlikely $\left(O\left(\epsilon_n^k\right)\right)$ that a future opponent will observe only interactions in which the agent played b because she observed the signal $m = k$. Thus the indirect gain of playing b when observing $m = k$ (which is $O\left(\epsilon_n\right)$) strictly outweighs the indirect loss (which is $O\left(\epsilon_n^k\right)$) and thus b is the unique best reply when an agent observes $m = k$. Hence $\left(S_n^N, \sigma_n, \theta_n\right)$ cannot be a Nash equilibrium if there is a strategy $s_n \in S_n^N$ such that $\left(s_n\right)_k (b) < 1$.

Case II: Suppose that $\pi(a, a) > \pi(b, b)$. We wish to show that a is a robust strictly perfect equilibrium action. Let $\left(S^C, \lambda\right)$ be an arbitrary distribution of commitments. Let $\bar{\beta} \in (0, 1)$ be the probability of action b in the unique mixed equilibrium of the underlying game G. Let $(\lambda|m) \in \Delta\left(S^C\right)$ be the posterior distribution of the partner's strategy, conditional on the partner following a commitment strategy, and the agent observing signal m about the partner, in a population in which everyone observes the signal $m = 0$ (which is the relevant case since we need $\theta_n \to_{n\to\infty} \theta^* \equiv \nu_\alpha$ in order for a to be a strictly perfect equilibrium action). Formally (by using Bayes' rule):

$$(\lambda|m)(s) = \frac{\lambda(s) \cdot \nu_{s_0}(m)}{\sum_{s \in S^C} \lambda(s) \cdot \nu_{s_0}(m)}.$$

Let $\beta_C(m)$ be the posterior probability that a random partner plays b conditional on (1) the agent observing signal m about the partner, (2) the partner following a commitment strategy, and (3) the partner observing signal 0 about the agent. Formally:

$$\beta_C(m) = \sum_{s \in S^C} (\lambda|m)(s) \cdot s_0(b).$$

It is straightforward to see that $\beta_C(m)$ is weakly increasing in m provided that ϵ is sufficiently small. (Note that if ϵ is very small then $s_0(b)$ is very close to the average probability that strategy s plays b.) Let $s^{\hat{m}}$ be the strategy that plays action a iff $m < \hat{m}$, i.e., $s_m^{\hat{m}}(a) = 1$ if $m < \bar{m}$, and $s_m^{\hat{m}}(a) = 0$ if $m \geq \hat{m}$.

Remark 11. In order to shorten the remaining proof, we take a simplifying assumption that for each m, $\beta_C(k) \neq \bar{\beta}$. The knife-edge cases in which $\beta_C(m) = \bar{\beta}$ for some $m \in M$ complicates the proof, and makes it substantially longer, which we felt is not justified given that the result is not the main focus of the paper.

To complete the proof (under the above simplifying assumption) we consider three exhaustive and mutually exclusive cases:

1. $\beta_C(k) < \bar{\beta}$. This implies that the steady state $(\{a\}, 1_a, \theta \equiv \nu_\alpha)$, where θ is any consistent behavior, is a Nash equilibrium of the perturbed environment $\left((G, k), \left(S^C, \lambda\right), \epsilon\right)$ for any sufficiently small $\epsilon > 0$.

2. $\beta_C(1) < \bar{\beta} \leq \beta_C(k)$. Let $\bar{m} > 1$ be the minimal signal m such that $\bar{\beta} < \beta_C(\bar{m})$. Then the steady state $(\{s^{\bar{m}}\}, 1_{s^{\bar{m}}}, \theta)$, where θ is any consistent signal profile in which the normal agents are observed to always play action a with a high probability of $\left(\theta_{s^{\bar{m}}}(0) > 1 - O(\epsilon)\right)$, is a Nash equilibrium of $\left((G, k), \left(S^C, \lambda\right), \epsilon\right)$

for any sufficiently small $\epsilon > 0$.

3. $\bar{\beta} < \beta_C(1)$. Let s^1 (s^2) be the strategy that induces an agent to play b iff $m \geq 1$ ($m \geq 2$). For each $q \in \left(0, \frac{1}{k}\right)$, consider the steady state $\left(\left\{s^1, s^2\right\}, (q, 1-q), \theta\right)$ of the perturbed environment $\left((G,k), \left(S^C, \lambda\right), \epsilon\right)$, where θ is a consistent signal profile in which the normal agents play a with an average probability of $1 - O(\epsilon)$ (such a consistent behavior exists due to the same arguments for the existence of a consistent behavior in which players cooperate with a probability of $1 - O(\epsilon)$ in the proof of Theorem 2).
Let μ_q be the posterior probability that a random partner is going to play b conditional on (1) the agent observing signal $m = 1$ about the partner, and (2) the partner observing signal $m = 0$ about the agent. Observe that (for a sufficiently small ϵ): (1) $\mu_0 = \beta_C(1) + O(\epsilon) > \bar{\beta}$, (2) μ_q is decreasing in q, and (3) $lim_{q \to \frac{1}{k}} \mu_q = O(\epsilon)$ (this is because each interaction in which a committed agent plays action b induces $O\left(\frac{1}{1-k \cdot q}\right)$ interactions in which normal agents play action b, as discussed in detail in the proof of Theorem 2 (see, in particular, Eq. (15)).
This implies that for every sufficiently small ϵ there is a value of q_ϵ such that $\mu_{q_\epsilon} = \bar{\beta}$ (and that this value converges to $q_0 \in \left(0, \frac{1}{k}\right)$ as ϵ converges to zero. In the steady state $\left(\left\{s^1, s^2\right\}, (q_\epsilon, 1-q_\epsilon), \theta\right)$ both actions a and b are best replies conditional on observing signal $m = 1$, while action a (b) is the unique best reply when observing $m = 0$ ($m > 1$). This implies that $\left(\left\{s^1, s^2\right\}, (q_\epsilon, 1-q_\epsilon), \theta\right)$ is a Nash equilibrium of[38] $\left((G,k), \left(S^C, \lambda\right), \epsilon\right)$.

In all three cases we have characterized a converging sequence of Nash equilibria of the perturbed environments $\left((G,k), \left(S^C, \lambda\right), \epsilon_n\right)$ in which all the normal agents play action a with an average probability of $1 - O(\epsilon)$, which implies that action a is strictly perfect.

D.4 Proof of Proposition 4 (Defection is Robust Strictly Perfect)

Let $\zeta = \left(\mathcal{S}^{\mathcal{C}}, \lambda\right)$ be a distribution of commitments. Let $s_d \equiv d$ be the strategy that always defects. Let $(\{s_d\}, \theta_n)$ be a steady state of the perturbed environment $\left((G,k), \left(S^C, \lambda\right), \epsilon_n\right)$. The fact that the behavior of the normal agent is independent of the observed signal implies that the steady state $(\{s_d\}, \theta_n)$ is robust and that $(\theta_n)_{s_d}(k) = 1$ (i.e., a player who follows strategy s_d is always observed to defect in all k interactions). Consider a deviating agent (Alice) who follows any strategy $s \neq s_d$. We show that Alice is strictly outperformed in any post-deviation steady state.

The facts that $s \neq s_d$ and that all signals are observed with positive probability in any perturbed environment imply that Alice cooperates with an average probability of $\alpha > 0$. We now compare the payoff of Alice to the payoff of an incumbent (Bob) who follows s_d. Alice obtains a direct loss of at least $\alpha \cdot \min(g, l)$ due to cooperating with probability α. The maximal indirect benefit that she might achieve due to these cooperations (by inducing committed agents to cooperate against her with higher probability relative to their cooperation probability against Bob) is $\epsilon_n \cdot k \cdot \alpha \cdot (l+1)$ because there are ϵ_n committed agents, each of whom observes Alice cooperate at least once in the k sampled actions with a probability of at most $k \cdot \alpha$, and each committed partner can yield Alice a benefit of at most $l+1$ by cooperating when the partner observes $m \geq 1$. If ϵ_n is sufficiently small ($\epsilon_n < \frac{1}{k \cdot (l+1)}$), then the direct loss is larger than the indirect maximal benefit ($\alpha > \epsilon_n \cdot k \cdot \alpha \cdot (l+1)$).

[38] We have abstracted away from a technical issue (which is formally investigated in the analogous arguments in the proof of Theorem 2). Specifically, we implicitly assumed that the probability that a random player defects conditional on both players observing $m = 1$ (denoted by the parameter $\chi \equiv \chi_{q_\epsilon}$ at the end of the proof of Theorem 2) is greater than μ_{q_ϵ}. Some distribution of commitment strategies might induce a situation in which $\chi_{q_\epsilon} < \mu_{q_\epsilon}$. In these cases, one needs to adapt the argument above by having the steady state $(\{s^{q_\epsilon}\}, 1_{s^{q_\epsilon}}, \theta)$, where s^{q_ϵ} is the strategy that plays b with probability q_ϵ when observing $m = 1$, play a for sure when observing $m = 0$, and play b for sure when observing $m > 1$.

This implies that $(\{s_d\}, \theta_n)$ is a (strict) Nash equilibrium in any environment with $\epsilon_n < \frac{1}{k \cdot (l+1)}$, which proves defection is a strictly perfect equilibrium action.

D.5 Proof of Theorem 1 (Defection is the Unique Equilibrium in Offensive PDs)

Let $(S^*, \sigma^*, \theta^*)$ be a regular perfect equilibrium. That is, there exists a regular distribution of commitments $\left(\mathcal{S}^{\mathcal{C}}, \lambda\right)$, a converging sequence $(\epsilon_n)_n \to 0$, and a converging sequence of steady states $\left(S_n^N, \sigma_n, \theta_n\right) \to (S^*, \sigma^*, \theta^*)$, such that for each n the state $\left(S_n^N, \sigma_n, \theta_n\right)$ is a Nash equilibrium of $\left((G,k), \left(S^C, \lambda\right), \epsilon_n\right)$. We assume to the contrary that $S^* \neq \{d\}$.

Recall that any signal $m \in M = \{0, ..., k\}$ is observed with positive probability in any perturbed environment. Given a state $\left(S_n^N, \sigma_n, \theta_n\right)$, an environment $\left((G,k), \left(S^C, \lambda\right), \epsilon_n\right)$, a signal $m \in M$, and a strategy $s \in S_n^N$, let $q(m,s)$ denote the probability that a randomly drawn partner of a player defects, conditional on the player following strategy s and observing signal m about the partner.

We say that a strategy is "defector-favoring" if the strategy is to defect against partners who are likely to cooperate, and to cooperate against partners who are likely to defect. Specifically, a strategy is defector-favoring if there is some threshold such that the strategy is to cooperate (defect) when the partner's conditional probability of defecting is above (below) this threshold. Formally:

Definition 18. Strategy $s \in S_n^N$ is *defector-favoring*, given state $\left(S_n^N, \sigma_n, \theta_n\right)$ and environment $\left((G,k), \left(S^C, \lambda\right), \epsilon_n\right)$, if there is some $\bar{q} \in [0,1]$ such that, for each $m \in M$, $q(m,s) > \bar{q} \Rightarrow s_m(d) = 0$, and $q(m,s) < \bar{q} \Rightarrow s_m(d) = 1$.

The rest of the proof consists of the following four steps.

First, we show that all normal strategies are defector-favoring. Assume to the contrary that there is a strategy $s \in S_n^N$ that is not defector-favoring. Let s' be a defector-favoring strategy that has the same average defection probability as s in the post-deviation steady state. The fact that both strategies prescribe defection with the same average probability implies that they induce the same behavior from the partners (since these partners observe identical distributions of signals when facing s and when facing s'), and hence $q(m,s) = q(m,s')$. Agents who follow strategy s' defect more often against partners who are more likely to cooperate relative to strategy s. Since the underlying game is offensive this implies that strategy s' strictly outperforms strategy s, which contradicts that $\left(S_n^N, \sigma_n, \theta_n\right)$ is a Nash equilibrium.

Second, we show that all the normal strategies lead agents to defect with the same average probability in $\left(S_n^N, \sigma_n, \theta_n\right)$. Assume to the contrary that there are strategies $s, s' \in S_n^N$ such that agents following the former strategy have a higher average probability of defection, i.e., $\alpha(\theta_s)(d) > \alpha(\theta_{s'})(d)$. Let $\beta = \alpha(\theta_s)(d) - \alpha(\theta_{s'})(d)$. Note that agents who follow strategy s have a strictly higher payoff than agents who follow s' when being matched with normal partners. This is because strategy s yields: (1) a strictly higher direct payoff of at least $\beta \cdot l$ due to playing more often the dominant action d, and (2) a weakly higher payoff against normal agents, because the fact that agents who follow it defect more often and all normal agents follow defector-favoring strategies implies that normal partners defect with a weakly smaller probability when being matched with agents who follow strategy s (relative to s'). We also need to consider what happens when normal agents are matched with committed agents. The maximal indirect gain that followers of strategy s' have relative to followers of strategy s, due to inducing a higher probability of cooperation from committed partners, is at most $\epsilon_n \cdot (l+1) \cdot k \cdot \beta$. This implies that if $\epsilon_n < \frac{l}{(l+1) \cdot k}$, then followers of strategy s have a strictly higher payoff than followers of s', which contradicts that $\left(S_n^N, \sigma_n, \theta_n\right)$ is a Nash equilibrium.

Third, we argue that for any normal agent it is the case that the probability that the partner defects conditional on the agent observing signal $m = k$ is weakly larger than the probability that the partner defects conditional on the agent observing any signal $m < k$. To see why, note that the regularity of the set of commitments implies that not all commitment strategies have the same defection probabilities, and thus the signal about the partner yields some information about the partner's probability of defecting. The previous step shows that all normal agents defect with the same probability, which implies that they induce the same signal distribution, and thus they induce the same behavior from all partners. Combining this fact with the fact that not all commitment strategies have the same defection probability implies (for a sufficiently small ϵ_n) that if a player observes a signal that includes only defections, then the partner is more likely to have a higher average defection probability against normal agents (i.e., $q(m,s) < q(k,s)$ for any normal strategy s and any $m < k$).

Thus, any normal agent (who follows a defector-favoring strategy due to the first step) defects with a weakly higher probability after observing signal $m = k$. This implies that if ϵ_n is sufficiently small, then a deviator who always defects outperforms the incumbents. The deviator achieves a direct higher payoff by defecting more often, as well as a weakly higher indirect gain by inducing the incumbents to cooperate more often.

D.6 Proof of Theorem 2 (Cooperation Is Strictly Perfect in Defensive PDs)

Part 1: Let $(S^*, \sigma^*, \theta^* \equiv 0)$ be a perfect equilibrium. This implies that there exist a distribution of commitments $\left(S^C, \lambda\right)$, a converging sequence of strictly positive commitment levels $\epsilon_n \to_{n\to\infty} 0$, and a converging sequence of steady states $\left(S_n^N, \sigma_n, \theta_n\right) \to_{n\to\infty} (S^*, \sigma^*, \theta^*)$, such that for each n the state $\left(S_n^N, \sigma_n, \theta_n\right)$ is a Nash equilibrium of the perturbed environment $\left((G,k), \left(S^C, \lambda\right), \epsilon_n\right)$. The fact that the equilibrium induces full cooperation (in the limit when $\epsilon_n \to_{n\to\infty} 0$) implies that all normal agents must cooperate when they observe no defections, i.e., $s_0(c) = 1$ for each $s \in S^*$.

Next we show that $s_1(d) > 0$ for some $s \in S^*$. Assume to the contrary that $s_1(d) = 0$ for every $s \in S^*$. This implies that for any $\delta > 0$, if n is sufficiently large then $\sum_{s \in S_n^N} \sigma_n(s) \cdot s_1(d) < \delta$. It follows that if a deviator (Alice) who follows a strategy s' defects with a small probability of $\alpha << 1$ when observing no defections (i.e., $s'_0(d) = \alpha$), then she outperforms the incumbents. To see this note that since she occasionally defects when observing $m = 0$ she obtains a direct gain of at least $\alpha \cdot g \cdot \Pr(m = 0)$, where $\Pr(m = 0)$ is the probability of observing $m = 0$ given the steady state $\left(S_n^N, \sigma_n, \theta_n\right)$. The probability that a partner observes her defecting twice or more is $O\left(\alpha^2\right)$. This implies that her indirect loss from these defections is at most $\left(O\left(\alpha^2\right) + O(\alpha) \cdot O(\delta + \epsilon_n)\right) \cdot (1 + l)$ and, thus, for sufficiently small values $\alpha, \delta, \epsilon_n > 0$, Alice strictly outperforms the incumbents.

We now show that $s_m(d) = 1$ for all $s \in S^*$ and all $m \geq 2$. The fact that $\theta^* \equiv 0$ implies that for a sufficiently large n, all normal agents cooperate with an average probability very close to one and, thus, the average probability of defection by an agent who follows a strategy $s \in S \cup S^C$ is very close to $s_0(d)$. Hence the distribution of signals induced by such an agent is very close to $\nu_{s_0(d)}$. Recall that we assume that the distribution of commitments contains at least one strategy s with $s_0(d) > 0$. This implies that the posterior probability that the partner is going to defect is strictly increasing in the signal m that the agent observes about the partner. Note that the direct gain from defecting is strictly increasing in the probability that the partner defects as well (due to the game being defensive), while the indirect influence of defection (on the behavior of future partners who may observe the current defection) is independent of the partner's play. From the previous paragraph we know that defection is a best reply conditional on an agent observing $m = 1$. This implies that defection must be the unique best reply when an agent observes at least two defections (i.e., when $m \geq 2$).

It remains to show that there is a normal incumbent strategy to cooperate with positive probability after observing a single defection, i.e., $s_1(d) < 1$ for some $s \in S^*$. Assume to the contrary that $s_1(d) = 1$ for every $s \in S^*$. Let r_n denote the average probability that a normal agent defects after observing $m \geq 1$. Since $\left(S_n^N, \sigma_n, \theta_n\right) \to_{n\to\infty} \left(S^*, \sigma^*, \theta^*\right)$, the assumption that $s_1(d) = 1$ for all $s \in S^*$ implies that $r_n > 0.6$ for a sufficiently large n. Let $Pr\left(m \geq 1 | S_n^N\right)$ denote the probability of observing $m \geq 1$ conditional on being matched with a normal partner. Note that the assumption that $\hat{s}_0(d) > 0$ for some committed strategy $\hat{s}$ and the assumption that $s_1(d) > 0$ for some normal strategy together imply that $Pr\left(m \geq 1 | S_n^N\right) > 0$. Note that $\theta^* \equiv c$ implies that $\lim_{n\to\infty}\left(Pr\left(m = 1 | S_n^N\right)\right) = 0$. Hence $Pr\left(m = 1 | S_n^N\right)$ is $O(\epsilon_n)$. We can calculate $Pr\left(m \geq 1 | S_n^N\right)$ as follows:

$$Pr\left(m \geq 1 | S_n^N\right) = k \cdot \left((1-\epsilon_n) \cdot r_n \cdot Pr\left(m \geq 1 | S_n^N\right) + \epsilon_n \cdot \lambda(\hat{s}) \cdot (\hat{s}_0(d) + O(\epsilon_n))\right) - O\left(\epsilon_n^2\right) - O\left(\left(Pr\left(m \geq 1 | S_n^N\right)\right)^2\right).$$

The reason for this equation is as follows. The observed signal induced by a normal agent (Bob) describes his actions in k interactions. In each of these interactions Bob's partner was normal with a probability of $1-\epsilon_n$, and was committed with a probability of ϵ_n. If Bob's partner in an interaction was normal then she defected with a probability of r_n when she observed $m \geq 1$ (which happened with a probability of $Pr\left(m \geq 1 | S_n^N\right)$). If Bob's partner in an interaction was committed then she followed strategy $\hat{s}$ with a probability of $\lambda(\hat{s})$ and defected with a probability of $\hat{s}_0(d) + O(\epsilon_n)$ (as argued above, the average defection probability of an agent following strategy s should be close to $s_0(d)$). Finally, the terms $-O\left(\epsilon_n^2\right) - O\left(\left(Pr\left(m \geq 1 | S_n^N\right)\right)^2\right)$ are subtracted to avoid "double-counting" cases in which Bob has defected more than once. Rearranging and simplifying the above equation by using the fact that $\left(Pr\left(m \geq 1 | S_n^N\right)\right)^2$ is $O\left(\epsilon_n^2\right)$ yields

$$(1 - k \cdot (1-\epsilon_n) \cdot r_n) \cdot Pr\left(m \geq 1 | S_n^N\right) = k \cdot (\epsilon_n \cdot \lambda(\hat{s}) \cdot \hat{s}_0(d)).$$

Then use $r_n > 0.6$ to infer that the LHS is negative. This contradicts the fact that the RHS is positive.

Part 2: Recall that s^1 (s^2) is the strategy that induces an agent to defect iff the agent observes $m \geq 1$ ($m \geq 2$). Let $0 < q < \frac{1}{k \cdot (l+1)}$ be a probability that will be defined later. Let s^q be the strategy that induces an agent to defect with a probability of q iff the agent observes $m = 1$, to defect for sure if she observes $m \geq 2$, and to cooperate for sure if she observes $m = 0$. Let $\left(S^C, \lambda\right)$ be an arbitrary distribution of commitments. We will show that there exist a converging sequence of commitment levels $\epsilon_n \to 0$ and converging sequences of steady states

$$\psi_n \equiv \left(\left\{s^1, s^2\right\}, \sigma_n = (q_n, 1-q_n), \theta_n\right) \to_{n\to\infty} \psi^* \equiv \left(\left\{s^1, s^2\right\}, (q, 1-q), \theta \equiv 0\right),$$

and

$$\psi'_n \equiv \left(\left\{s^{q_n}\right\}, \theta'_n\right) \to_{n\to\infty} \psi'^* \equiv \left(\left\{s^q\right\}, 1_{s^q}, \theta' \equiv 0\right),$$

such that either (1) for each n the steady state ψ_n is a Nash equilibrium of $\left((G,k), \left(S^C, \lambda\right), \epsilon_n\right)$, or (2) for each n the steady state ψ'_n is a Nash equilibrium of $\left((G,k), \left(S^C, \lambda\right), \epsilon_n\right)$.

Fix an $n \geq 1$ such that ϵ_n is sufficiently small. (Exactly what counts as sufficiently small will become clear below.) In what follows, we calculate a number of probabilities while relying on the fact that $\epsilon_n << 1$. Thus we neglect terms of $O(\epsilon_n)$ (resp., $O\left(\epsilon_n^2\right)$) when the leading term is $O(1)$ (resp., $O(\epsilon_n)$). The calculations give the same results for ψ_n as for ψ'_n. Since we are looking for consistent signal profiles θ_n and θ'_n such that $\theta_n \to_{n\to\infty} \theta \equiv 0$ and $\theta'_n \to_{n\to\infty} \theta' \equiv 0$, we assume that $(\theta_n)_{s_i}(0) = 1 - O(\epsilon_n)$ for each $s_i, s_j \in \left\{s^1, s^2\right\}$ in ψ_n and assume that $(\theta'_n)_{s^q}(0) = 1 - O(\epsilon_n)$ in ψ'_n.

We begin by confirming that indeed there exist consistent signal profiles θ_n and θ'_n in which the normal agents almost always cooperate, and that, moreover, the steady states ψ_n and ψ'_n satisfy the robustness property. Consider a perturbed signal profile $\theta \in O_{(S_n^N \cup S^C)}$. Recall that $\alpha_{\sigma_n}(\theta)(d)$ is the (σ_n-weighted) average of the distributions of actions that induce signals distributed according to the signal profile θ for the normal agents, i.e.,

$$\alpha_{\sigma_n}(\theta)(d) = q_n \cdot \alpha(\theta_{s^1})(d) + (1-q_n) \cdot \alpha(\theta_{s^2})(d) \quad (\alpha_{\sigma_n}(\theta)(d) = \alpha(\theta_{s^q})(d)).$$

The (possibly inconsistent) "old" perturbed signal profile θ and the strategy distribution of the incumbents jointly induce a "new" signal profile $f_{(1-\epsilon_n)\cdot\sigma_n+\epsilon_n\cdot\lambda}(\theta)$. The average defection probability of a normal agent in this "new" signal profile is bounded by the following inequality:

$$\alpha_{\sigma_n}\left(f_{(1-\epsilon_n)\cdot\sigma_n+\epsilon_n\cdot\lambda}(\theta)\right)(d) < (1-\epsilon_n)\cdot\left(q_n\cdot k\cdot\alpha_{\sigma_n}(\theta)(d) + \binom{k}{2}\cdot\left(\alpha_{\sigma_n}(\theta)(d)\right)^2\right) + \epsilon_n. \tag{14}$$

This is so because a normal agent, when being matched with a normal partner (which happens with a probability of $(1-\epsilon_n)$) defects with an average probability of q_n when she observes a single defection (which happens with a probability strictly less than $k\cdot\alpha_{\sigma_n}(\theta)$), and defects for sure when she observes at least two defections (which happens with a probability strictly less than $\binom{k}{2}\cdot(\alpha_{\sigma_n}(\theta))^2$). Consider the parabolic equation, which is based on substituting

$$x = \alpha_{\sigma_n}(\theta)(d) = \alpha_{\sigma_n}\left(f_{(1-\epsilon_n)\cdot\sigma_n+\epsilon_n\cdot\lambda}(\theta)\right)(d)$$

in (14), and changing the inequality into an equality:

$$x = (1-\epsilon_n)\cdot\left(q_n\cdot k\cdot x + \binom{k}{2}\cdot x^2\right) + \epsilon_n \Leftrightarrow 0 = \binom{k}{2}\cdot x^2 - (1-(1-\epsilon_n)\cdot q_n\cdot k)\cdot x + \epsilon_n.$$

Recall that a parabolic equation $A\cdot x^2 - B\cdot x + c = 0$ with $A,B,C>0$ and $C << A,B$ has two positive solutions, the smaller of which is

$$x_1 = \frac{B-\sqrt{B^2-4\cdot A\cdot C}}{2\cdot A} \approx \frac{B-\sqrt{B^2-2\cdot B\cdot\frac{2\cdot A\cdot C}{B}+\left(\frac{2\cdot A\cdot C}{B}\right)^2}}{2\cdot A} = \frac{B-\left(B-\frac{2\cdot A\cdot C}{B}\right)}{2\cdot A} = \frac{\frac{2\cdot A\cdot C}{B}}{2\cdot A} = \frac{C}{B} = \frac{\epsilon_n}{1-(1-\epsilon_n)\cdot q_n\cdot k} = \kappa_n\cdot\epsilon_n,$$

where the penultimate equality is derived by substituting $C=\epsilon_n$ and $B = 1-(1-\epsilon_n)\cdot q_n\cdot k$, and the last equality is derived by defining $\kappa_n = \frac{1}{1-(1-\epsilon_n)\cdot q_n\cdot k}$. Let $\kappa = sup_n\kappa_n < \infty$. The upper bound κ is finite due to the fact that $q_n \to q$, $k\cdot q < \frac{1}{l+1}$ and $\epsilon_n \to \epsilon$. The definition of $x_1 = \kappa_n\cdot\epsilon_n$ implies that

$$\alpha_{\sigma_n}(\theta)(d) \le \kappa_n\cdot\epsilon_n \;\Rightarrow\; \alpha_{\sigma_n}\left(f_{(1-\epsilon_n)\cdot\sigma_n+\epsilon_n\cdot\lambda}(\theta)\right)(d) < \kappa_n\cdot\epsilon_n,$$

which immediately implies the robustness property of the action c :

$$\alpha_{\sigma_n}(\theta)(c) \ge 1-\kappa_n\cdot\epsilon_n \;\Rightarrow\; \alpha_{\sigma_n}\left(f_{(1-\epsilon_n)\cdot\sigma_n+\epsilon_n\cdot\lambda}(\theta)\right)(c) > 1-\kappa_n\cdot\epsilon_n.$$

Let $O_{(\{s^1,s^2\}\cup S_C,x_1)}$ $(O_{(s_{q_n}\cup S_C,x_1)})$ be the set of signal profiles θ defined over $\{s^1,s^2\}\cup S_C$ $(s_{q_n}\cup S_C)$

and satisfying $\alpha_{\sigma_n}(\theta)(d) \leq x_1$. Observe that $O_{(\{s^1,s^2\}\cup S_C,x_1)}(O_{(s_{q_n}\cup S_C,x_1)})$ is a convex and compact subset of a Euclidean space, and that the mapping $f_{(1-\epsilon_n)\cdot\sigma_n+\epsilon_n\cdot\lambda}(\theta)$ is continuous. Brouwer's fixed-point theorem implies that the mapping $f_{(1-\epsilon_n)\cdot\sigma_n+\epsilon_n\cdot\lambda}(\theta)$ has a fixed point θ_n (θ'_n) satisfying $\alpha_{\sigma_n}(\theta_n)(d) < x_1 = O(\epsilon_n)$ ($\alpha_{\sigma_n}(\theta'_n)(d) \leq x_1 = O(\epsilon_n)$), which is a consistent signal profile in which the normal agents almost always cooperate.

For each incumbent strategy s, let $Pr(m=1|s)$ ($Pr(m\geq 2|s)$) denote the probability of observing exactly one defection (at least two defections) conditional on the partner following strategy s. Let $Pr(m=1)$ and $Pr(m\geq 2)$ be the corresponding unconditional probabilities.

The assumption that $\theta_n \to_{n\to\infty} \theta \equiv 0$ and $\theta'_n \to_{n\to\infty} \theta' \equiv 0$ implies that agents are very likely to observe the signal $m=0$ (i.e., zero defections) when being matched with a random partner. Formally:

$$Pr(m=0) = (1-O(\epsilon_n))^k = 1-O(\epsilon_n).$$

The conditional probabilities of observing $m=0$, $m=1$, and $m\geq 2$, for all $s \in S_n^N \cup S^C$, are

$$Pr(m=0|s) = (s_0(c))^k + O(\epsilon_n),$$

$$Pr(m=1|s) = k\cdot s_0(d)\cdot(s_0(c))^{k-1} + O(\epsilon_n),$$

$$Pr(m\geq 2|s) = 1 - Pr(m=0|s) - Pr(m=1|s).$$

Let $S_n^N = \{s^1, s^2\}$ in ψ_n and $S_n^N = \{s^{q_n}\}$ in ψ'_n. Given signal m, let $Pr\left(m|S_n^N\right)$ denote the probability of observing signal m, conditional on the partner following a normal strategy. Specifically, in the heterogeneous state ψ_n (with two normal strategies), this conditional probability is given by

$$Pr\left(m|S_n^N\right) = q\cdot Pr\left(m|s^1\right) + (1-q)\cdot Pr\left(m|s^2\right).$$

Furthermore, it follows (from the expressions for $Pr(m=0|s)$, $Pr(m=1|s)$, and $Pr(m\geq 2|s)$) that

$$Pr\left(m=0|S_n^N\right) = 1-O(\epsilon_n), \quad Pr\left(m=1|S_n^N\right) = O(\epsilon_n) \quad Pr\left(m\geq 2|S_n^N\right) = O\left(\epsilon_n^2\right).$$

Next we calculate the probability that a normal agent (Alice) generates a signal that contains a single defection. This happens with probability one if exactly one of the k interactions sampled from Alice's past was such that Alice observed her partner in that interaction to have defected at least twice (which implies that her partner is most likely to have been a committed agent). This happens with probability q_n if exactly one of the k interactions sampled from Alice's past was such that Alice observed her partner (who might have been either a committed or a normal agent) to have defected exactly once:

$$\begin{aligned} Pr\left(m=1|S_n^N\right) &= k\cdot\sum_{s\in S^C}\epsilon_n\cdot\lambda(s)\cdot(Pr(m\geq 2|s) + q_n\cdot Pr(m=1|s)). \\ &\quad +k\cdot(1-\epsilon_n)\cdot\left[q_n\cdot Pr\left(m=1|S_n^N\right) + Pr\left(m\geq 2|S_n^N\right)\right] \\ &\quad +O\left(\epsilon_n^2\right). \end{aligned}$$

The final term $O\left(\epsilon_n^2\right)$ comes from the very small probability of the partner observing a normal agent to defect twice. Since $Pr\left(m=1|S_n^N\right) = O(\epsilon_n)$ and $Pr\left(m\geq 2|S_n^N\right) = O\left(\epsilon_n^2\right)$, this can be simplified (neglecting $O\left(\epsilon_n^2\right)$)

and rearranged to obtain

$$Pr\left(m=1|S_n^N\right)=\frac{k\cdot\epsilon_n\cdot\sum_{s\in S^C}\lambda\left(s\right)\cdot\left(Pr\left(m\geq 2|s\right)+q_n\cdot Pr\left(m=1|s\right)\right)}{1-k\cdot q}, \tag{15}$$

which is well defined and $O\left(\epsilon_n\right)$ as long as $q_n<1/k$. We can now calculate the unconditional probabilities:

$$Pr\left(m=1\right)=\epsilon_n\cdot\sum_{s\in S^C}\lambda\left(s\right)\cdot Pr\left(m=1|s\right)+Pr\left(m=1|S_n^N\right)+O\left(\epsilon_n^2\right),$$

$$\begin{aligned}Pr\left(m\geq 2\right) &= \epsilon_n\cdot\sum_{s\in S^C}\lambda\left(s\right)\cdot Pr\left(m\geq 2|s\right)+\left(1-\epsilon_n\right)\cdot Pr\left(m\geq 2|S_n^N\right)\\ &= \epsilon_n\cdot\sum_{s\in S^C}\lambda\left(s\right)\cdot Pr\left(m\geq 2|s\right)+O\left(\epsilon_n^2\right).\end{aligned}$$

By using Bayes' rule we can calculate the conditional probability that the partner uses strategy $s\in S^C$ as a function of the observed signal:

$$\begin{aligned}Pr\left(s|m=0\right) &= \frac{\epsilon_n\cdot\lambda\left(s\right)\cdot Pr\left(m=0|s\right)}{Pr\left(m=0\right)},\\ Pr\left(s|m=1\right) &= \frac{\epsilon_n\cdot\lambda\left(s\right)\cdot Pr\left(m=1|s\right)}{Pr\left(m=1\right)},\\ Pr\left(s|m\geq 2\right) &= \frac{\epsilon_n\cdot\lambda\left(s\right)\cdot Pr\left(m\geq 2|s\right)}{Pr\left(m\geq 2\right)}.\end{aligned}$$

Note that

$$\sum_{s\in S^C}Pr\left(s|m=0\right) = \frac{\epsilon_n\cdot\sum_{s\in S^C}\lambda\left(s\right)\cdot\left(s_0\left(c\right)\right)^k}{1-O\left(\epsilon_n\right)}=O\left(\epsilon_n\right).$$

From Eq. (15) we have

$$\sum_{s\in S_n^N}\sigma\left(s\right)\cdot Pr\left(m=1|s\right) = Pr\left(m=1|S_n^N\right)=\frac{k\cdot\epsilon_n\cdot\sum_{s\in S^C}\lambda\left(s\right)\cdot\left(Pr\left(m\geq 2|s\right)+q\cdot Pr\left(m=1|s\right)\right)}{1-k\cdot q_n}.$$

We use this to obtain, by Bayes' rule,

$$\begin{aligned}\sum_{s\in S^C}Pr\left(s|m=1\right) &= \frac{\epsilon_n\cdot\sum_{s\in S^C}\lambda\left(s\right)\cdot Pr\left(m=1|s\right)}{\epsilon_n\cdot\sum_{s\in S^C}\lambda\left(s\right)\cdot Pr\left(m=1|s\right)+\frac{k\cdot\epsilon_n\cdot\sum_{s\in S^C}\lambda(s)\cdot\left(Pr(m\geq 2|s)+q_n\cdot Pr(m=1|s)\right)}{1-k\cdot q_n}+O\left(\epsilon_n^2\right)}\\ &= \frac{\sum_{s\in S^C}\lambda\left(s\right)\cdot Pr\left(m=1|s\right)}{\sum_{s\in S^C}\lambda\left(s\right)\cdot Pr\left(m=1|s\right)+\frac{k\cdot\sum_{s\in S^C}\lambda(s)\cdot\left(Pr(m\geq 2|s)+q_n\cdot Pr(m=1|s)\right)}{1-k\cdot q_n}+O\left(\epsilon_n^2\right)}.\end{aligned}$$

Note that the terms $\sum_{s\in S^C}\lambda\left(s\right)\cdot Pr\left(m=1|s\right)$ and $\sum_{s\in S^C}\lambda\left(s\right)\cdot\left(Pr\left(m\geq 2|s\right)\right)$ do not vanish as $\epsilon_n\to 0$. Moreover, we will see below (Eqs. (17) and (18)) that this implies that q_n also does not vanish as $\epsilon_n\to 0$. Together, these observations imply that there are numbers $a,b\in(0,1)$ such that, for all n, it is the case that

$$0<a<\sum_{s\in S^C}Pr\left(s|m=1\right)<b<1. \tag{16}$$

Furthermore

$$\sum_{s\in S^C} Pr\left(s|m\geq 2\right)=\frac{1}{1+\frac{\sum_{s\in S_n^N}\sigma(s)\cdot Pr(m\geq 2|s)}{\epsilon_n\cdot\sum_{s\in S^C}\lambda(s)\cdot Pr(m\geq 2|s)}}=\frac{1}{1+\frac{O(\epsilon_n^2)}{O(\epsilon_n)}}=\frac{1}{1+O\left(\epsilon_n\right)}.$$

Hence for a sufficiently large n, the more defections there are in the observed signal, the higher is the conditional probability that the partner is committed:

$$\sum_{s\in S^C} Pr\left(s|m=0\right)<\sum_{s\in S^C} Pr\left(s|m=1\right)<\sum_{s\in S^C} Pr\left(s|m\geq 2\right).$$

Let $Pr\left(S_n^N|m=1\right)=\sum_{s\in S_n^N} Pr\left(s|m=1\right)$ denote the conditional probability that the partner follows a normal strategy conditional on the agent observing signal $m=1$. Eq. (16) implies that there are numbers $a',b'\in(0,1)$ such that, for all n, it is the case that $0<a'<Pr\left(S_n^N|m=1\right)<b'<1$ (because $Pr\left(S_n^N|m=1\right)+\sum_{s\in S^C} Pr\left(s|m=1\right)=1$).

Let μ_n be the probability that a random partner defects conditional on a player observing signal $m=1$ about the partner, and conditional on the partner observing the signal $m=0$:

$$\mu_n=\sum_{s\in S^C} Pr\left(s|m=1\right)\cdot s_0\left(d\right)+O\left(\epsilon_n\right). \tag{17}$$

Eq. (17) defines μ_n as a strictly decreasing function of q_n. To see this, note that the term $s_0\left(d\right)$ does not depend on q_n, and in $Pr\left(s|m=1\right)=\frac{\epsilon_n\cdot\lambda(s)\cdot Pr(m=1|s)}{Pr(m=1)}$ the numerator does not depend on q_n, whereas the term $Pr\left(m=1\right)$ is increasing in q_n.

Next we calculate the value of q_n that balances the payoff of both actions after a player observes a single defection (neglecting terms of $O\left(\epsilon_n^2\right)$). The LHS of the following equation represents the player's direct gain from defecting when she observes a single defection, while the RHS represents the player's indirect loss induced by partners who defect as a result of observing these defections:

$$\Pr\left(m=1\right)\cdot\left(\mu_n\cdot l+\left(1-\mu_n\right)\cdot g\right)=\Pr\left(m=1\right)\cdot\left(k\cdot q\cdot\left(l+1\right)+O\left(\epsilon_n\right)\right)\Rightarrow q_n=\frac{\mu_n\cdot l+\left(1-\mu_n\right)\cdot g}{k\cdot\left(l+1\right)}+O\left(\epsilon_n\right). \tag{18}$$

Note that Eq. (18) defines q_n as a strictly increasing function of μ_n. This implies that there are unique values of q_n and μ_n satisfying $\frac{g}{k\cdot(l+1)}<q_n<\frac{l}{k\cdot(l+1)}<\frac{1}{k}$ and $0<\mu_n<1$, which jointly solve Eqs. (17) and (18). This pair of parameters balances the payoff of both actions when a player observes a signal $m=1$. Note that sequences of $\left(q_n\right)_n\to q$ and $\left(\mu_n\right)_n\to\mu$ converge to the values that solve the above equations when ignoring the terms that are $O\left(\epsilon_n\right)$.

Observe that defection is the unique best reply when a player observes at least two defections. The direct gain from defecting is larger than the LHS of Eq. (18), and the indirect loss is still given by the RHS of Eq. (18). The reason that the direct gain is larger is that normal partners almost never defect twice or more (the probability is $O\left(\epsilon_n^2\right)$), and thus the partner is most likely committed and will defect with a probability that is higher than μ_n (since μ_n also gives weight to normal strategies that are most likely to cooperate). More generally, note that given that the normal agents almost always cooperate, the average probability of defection of each agent who follows strategy s is $s_0\left(d\right)+O\left(\epsilon_n\right)$. This implies that for a sufficient small ϵ_n, the higher m is, the higher the partner's value $s_0\left(d\right)$ is likely to be. Hence the higher m is, the higher the probability is that the partner will defect against a normal agent. Thus the direct gain from defection is increasing in the signal m

that the normal agent observes about her partner. (A formal detailed proof of this statement is available upon request.)

Next, consider a deviator (Alice) who defects with a probability of $\alpha > 0$ after she observes $m = 0$. In what follows we calculate Alice's expected payoff as a function of α in any post-deviation stable state, neglecting terms of $O(\epsilon_n)$ throughout the calculation. Note that Alice's partner observes signal $m = 1$ with a probability of $k \cdot \alpha \cdot (1-\alpha)^{k-1}$, and observes signal $m \geq 2$ with a probability of $1 - (1-\alpha)^k - k \cdot \alpha \cdot (1-\alpha)^{k-1}$. This implies that the mean probability that a normal partner defects against a mutant is

$$h(\alpha) := \left(k \cdot \alpha \cdot (1-\alpha)^{k-1}\right) \cdot q + 1 - (1-\alpha)^k - k \cdot \alpha \cdot (1-\alpha)^{k-1} = 1 - (1-\alpha)^{k-1}(1 - \alpha + k \cdot \alpha \cdot (1-q)).$$

Thus the expected payoff of the mutant is

$$\begin{aligned} \pi(\alpha): &= (1-h(\alpha)) \cdot \alpha \cdot (1+g) + (1-h(\alpha)) \cdot (1-\alpha) - h(\alpha) \cdot (1-\alpha) \cdot l \\ &= 1 + \alpha \cdot g - h(\alpha) \cdot (1 + (1-\alpha) \cdot l + \alpha \cdot g). \end{aligned}$$

Direct numeric calculation of $\frac{\partial \pi(\alpha)}{\partial \alpha}$ reveals that $\pi(\alpha)$ is strictly decreasing in α for each $q > \frac{g}{k \cdot (l+1)}$. Thus any deviator with $\alpha > 0$ earns strictly less than the incumbents (who have $\alpha = 0$).

We have now shown that the best reply is c after observing $m = 0$ and d after observing $m \geq 2$. After observing $m = 1$ both c and d are best replies provided that q has the required value. That is, we know what the aggregate probability of defection after a player observes $m = 1$ has to be in equilibrium. However, we do not know whether mixing will occur at the individual level. We now turn to this question.

Let χ be the probability that a random partner defects conditional on both the agent and the partner observing a single defection (in the limit as $\epsilon_n \to 0$):

$$\chi = \lim_{n \to \infty} \left(\sum_{s \in S^C} Pr(s|m=1) \cdot s^1(d) + Pr\left(S_n^N | m = 1\right) \cdot q \right).$$

We conclude by showing that if $\chi > \mu$ ($\chi < \mu$), then ψ^* (ψ'^*) is a perfect equilibrium. This is so because if $\chi > \mu$ ($\chi < \mu$), then conditional on a normal agent observing a single defection, the partner is more (less) likely to defect the higher the probability with which the agent defects when she observes a single defection (because then it is more likely that the partner observes a single defection rather than only cooperation). This implies that when a player observes a single defection, the higher the agent's own defection probability is, the more profitable defection is (recall that the higher the probability is of the defection of the partner, the higher the direct gain from defection, whereas the indirect loss is independent of the partner's behavior). That is, an agent's payoff is a strictly convex (concave) function of the agent's defection probability conditional on him observing a single defection. This implies that a deviator who mixes on the individual level (i.e., defects with probabilities different from q) is outperformed when $\chi > \mu$ ($\chi < \mu$)).

Note that the normal agents are more likely to defect against a partner who is more likely to defect when she observes a single defection. This implies that when focusing only on normal partners, the induced level of χ is larger than the induced level of μ. It is only the committed agents who may induce the opposite inequality (namely, $\chi < \mu$). Thus, if in the limit as $\epsilon \to 0$ the equality $\chi = \mu$ holds, then it must be that for any positive small share of committed agents ϵ_n, it is the case that $\chi_n < \mu_n$, which implies by the argument above that the state ψ'_n is a Nash equilibrium.

Remark 12. The above argument shows that when $\chi < \mu$, each state ψ'_n is a *strictly* perfect equilibrium (any

deviator who follows a strategy different from s^{q_n} obtains a strictly lower payoff). In the opposite case of $\chi > \mu$, one can show that an agent who follows strategy s_i achieves a higher payoff than an agent who follows s_{-i}, conditional on the partner following s_i. This implies that the mixed equilibrium between the strategies of s^1 and x_2 is Hawk-Dove-like, and that the state ψ_n is evolutionarily stable (see Appendix B). This shows that cooperation is robust also to joint deviation of a small group of agents, and that it satisfies the refinement of evolutionary stability defined in Appendix B (namely, cooperation is a strictly perfect evolutionarily stable action).

D.7 Proof of Proposition 5 (Observing a Single Action)

Arguments and pieces of notation that are analogous to the ones used in the proof of Theorem 2 are presented in brief or skipped. Let $s^c \equiv c$ be the strategy that always cooperates. The same arguments as in Theorem 2 show that the only possible candidates for perfect equilibria that support full cooperation are steady states of the form $\psi = \left(\left\{s^1, s^c\right\}, (q, 1-q), \theta \equiv 0\right)$ or $\psi' = \left(\left\{s^q\right\}, 1_{s^q}, \theta' \equiv 0\right)$.

Consider a perturbed environment $\left(\left(G_{PD}, k\right), \left(S^C, \lambda\right), \epsilon\right)$ where $\epsilon > 0$ is sufficiently small. In what follows: (1) for the case of $g \leq \beta_{C,\lambda}$ we characterize a Nash equilibrium of this perturbed environment that is within a distance of $O(\epsilon)$ from either ψ or ψ', and (2) we show that no such Nash equilibrium exists for the case of $g > \beta_{C,\lambda}$.

Consider a steady state that is within a distance of $O(\epsilon)$ from either ψ or ψ'. The fact that the behavior in the steady state is close to always cooperating (i.e., to $\theta \equiv 0$) implies that the probability of observing $m = 1$ conditional on the partner following a commitment strategy $s \in S^C$ is:

$$Pr(m = 1|s) = s_0(d) + O(\epsilon).$$

Similarly, the probability of observing $m = 1$ conditional on the partner being normal is

$$Pr\left(m = 1|S_n^N\right) = q \cdot \left(\epsilon \cdot \sum_{s \in S^C} \lambda(s) \cdot s_0(d) + (1-\epsilon) \cdot Pr\left(m = 1|S_n^N\right)\right) + O\left(\epsilon^2\right) \Rightarrow$$

$$Pr\left(m = 1|S_n^N\right) = \frac{\epsilon \cdot q \cdot \sum_{s \in S^C} \lambda(s) \cdot s_0(d)}{1-q} + O\left(\epsilon^2\right).$$

By using Bayes' rule we can calculate the probability that the partner uses strategy $s \in S^C$ conditional on observing $m = 1$:

$$Pr(s|m=1) = \frac{\epsilon_n \cdot \lambda(s) \cdot Pr(m=1|s)}{Pr(m=1)} = \frac{\epsilon \cdot \lambda(s) \cdot s_0(d)}{\epsilon \cdot \left(\sum_{s \in S^C} \lambda(s) \cdot s_0(d) + \frac{q \cdot \sum_{s \in S^C} \lambda(s) \cdot s_0(d)}{1-q}\right)} + O(\epsilon) \Rightarrow$$

$$Pr(s|m=1) = \frac{(1-q) \cdot \lambda(s) \cdot s_0(d)}{\sum_{s \in S^C} \lambda(s) \cdot s_0(d)} + O(\epsilon).$$

Let μ be the probability that a random partner defects conditional on an agent observing signal $m = 1$ about the partner, and conditional on the partner observing the signal $m = 0$ about the agent. (Note that only committed partners defect with positive probability when observing $m = 0$.)

$$\mu = \sum_{s \in S^C} Pr(s|m=1) \cdot s_0(d) + O(\epsilon) = (1-q) \cdot \frac{\sum_{s \in S^C} \lambda(s) \cdot (s_0(d))^2}{\sum_{s \in S^C} \lambda(s) \cdot s_0(d)} + O(\epsilon) = (1-q) \cdot \beta_{(S^C,\lambda)} + O(\epsilon). \quad (19)$$

Next we calculate the value of q that balances the payoff of both actions after a player observes a single defection. The LHS of the following equation represents the player's direct gain from defecting when she observes a single defection, while the RHS represents the player's indirect loss induced by future partners who defect as a result of observing these defections:

$$\Pr(m=1)\cdot(\mu\cdot l+(1-\mu)\cdot g)\quad +O(\epsilon)= \quad \Pr(m=1)\cdot(q\cdot(l+1)+O(\epsilon))\Rightarrow \tag{20}$$

$$q=\frac{\mu\cdot l+(1-\mu)\cdot g}{l+1}+O(\epsilon)=\frac{g+\mu\cdot(l-g)}{l+1}+O(\epsilon). \tag{21}$$

Substituting (19) in (21) yields

$$q=\frac{g+(1-q)\cdot(l-g)\cdot\beta_{(S^C,\lambda)}}{l+1}+O(\epsilon)\Rightarrow q\cdot(l+1)=g+(1-q)\cdot(l-g)\cdot\beta_{(S^C,\lambda)}+O(\epsilon)$$

$$\Rightarrow q=\frac{g+(l-g)\cdot\beta_{(S^C,\lambda)}}{l+1+(l-g)\cdot\beta_{(S^C,\lambda)}}+O(\epsilon).$$

Consider a deviator (Alice) who always defects. Normal partners of Alice cooperate with a probability of $1-q$. This implies that Alice gets an expected payoff of $(1+g)\cdot(1-q)$, while the normal agents each get a payoff of $1+O(\epsilon)$. Alice is outperformed iff (neglecting terms of $O(\epsilon)$):

$$(1+g)\cdot(1-q)\leq 1\Leftrightarrow q\geq\frac{g}{1+g}\Leftrightarrow\frac{g+(l-g)\cdot\beta_{(S^C,\lambda)}}{l+1+(l-g)\cdot\beta_{(S^C,\lambda)}}\geq\frac{g}{1+g}$$

$$\Leftrightarrow(1+g)\cdot\left(g+(l-g)\cdot\beta_{(S^C,\lambda)}\right)\geq g\cdot\left(l+1+(l-g)\cdot\beta_{(S^C,\lambda)}\right)$$

$$\Leftrightarrow g^2+(l-g)\cdot\beta_{(S^C,\lambda)}+\geq g\cdot l\ \Leftrightarrow\ g\cdot(l-g)\leq(l-g)\cdot\beta_{(S^C,\lambda)}\ \Leftrightarrow\ g\leq\beta_{(S^C,\lambda)}.$$

Thus, the steady state can be a Nash equilibrium only if $g\leq\beta_{(S^C,\lambda)}$. It is relatively straightforward to show that if $g\leq\beta_{(S^C,\lambda)}$, then a deviator who defects with probability α when observing $m=0$ is outperformed. The remaining steps of the proof are as in the proof of part 2 of Theorem 2, and are omitted for brevity.

D.8 Proof of Theorem 3 (Observing Conflicts)

The proof of part 1(a) is analogous to Theorem 2 and is omitted for brevity. We now prove Part 1(b), i.e., that any mild game admits a strictly perfectly equilibrium action. Arguments and notations that are analogous to the proof of Theorem 2 are presented in brief. Let s^1 (s^2) be the strategy that instructs a player to defect if and only if she receives a signal containing one or more (two or more) conflicts. Consider the following candidate for a perfect equilibrium $\left(\left\{s^1,s^2\right\},(q,1-q),\theta^*=0\right)$. Here, the probability q will be determined such that both actions are best replies when an agent observes a single conflict.

Let $\left(\mathcal{S}^{\mathcal{C}},\lambda\right)$ be a distribution of commitments. We show that there exists a converging sequence of levels $\epsilon_n\to 0$, and converging sequences of steady states $\left(\left\{s^1,s^2\right\},(q_n,1-q_n),\theta_n\right)\to\left(\left\{s^1,s^2\right\},(q,1-q),\theta\equiv 0\right)$ and $\left(\left\{s^{q_n}\right\},1_{s^{q_n}},\theta'_n\right)\to\left(\left\{s^q\right\},1_{s^q},\theta'\equiv 0\right)$ such that either (1) each steady state $\psi_n\equiv\left(\left\{s^1,s^2\right\},\sigma_n\equiv(q_n,1-q_n),\theta_n\right)$ is a Nash equilibrium of $\left((G,k),\left(S^C,\lambda\right),\epsilon_n\right)$, or (2) each steady state $\psi'_n\equiv\left(\left\{s^{q_n}\right\},\sigma'_n\equiv 1_{s^{q_n}},\theta'_n\right)$ is a Nash equilibrium of $\left((G,k),\left(S^C,\lambda\right),\epsilon_n\right)$.

Fix $n\geq 1$. Assume that ϵ_n is sufficiently small. We calculate the probability $Pr\left(m=1|S_n^N\right)$ that a normal agent (Alice) induces a signal $m=1$. Since we focus on the steady states in which the incumbents defect very rarely (i.e., θ_n and θ'_n converge to $\theta^*\equiv 0$), we can assume that $Pr\left(m=1|S_n^N\right)$ is $O(\epsilon_n)$. (The proof of the

existence and of the robustness of consistent signal profiles in which the normal agents almost always cooperate in mild PDs is analogous to the argument presented in the proof of Theorem 2, and is omitted for brevity). Alice may be involved in a conflict if one of her k partners is committed, which happens with a probability of $O(\epsilon_n)$. If all of the k partners are normal, then at each interaction both Alice and her partner defect with a probability of $Pr\left(m=1|S_n^N\right)$, which implies that the probability of a conflict is $2\cdot Pr\left(m=1|S_n^N\right)-\left(Pr\left(m=1|S_n^N\right)^2\right)$. Therefore:

$$Pr\left(m=1|S_n^N\right)=k\cdot\left(O(\epsilon_n)+2\cdot q_n\cdot Pr\left(m=1|S_n^N\right)-O\left(\left(Pr\left(m=1|S_n^N\right)\right)^2\right)\right).$$

Solving this equation, while neglecting terms that are $O\left(\epsilon_n^2\right)$ (including $Pr\left(m=1|S_n^N\right)^2$), yields

$$Pr\left(m=1|S_n^N\right)=\frac{k\cdot O(\epsilon_n)}{1-2\cdot k\cdot q_n}, \tag{22}$$

which is well defined and $O(\epsilon_n)$ as long as $q_n<\frac{1}{2\cdot k}$. Note that as q_n approaches $\frac{1}{2\cdot k}$, the value of $Pr\left(m=1|S_n^N\right)$ "explodes" (becomes arbitrarily larger than terms that are $O(\epsilon_n)$).

By Bayes' rule we can calculate the conditional probability $Pr(s|m=1)$ of being matched with each strategy $s\in S^C$ (same calculations as detailed in the proof of Theorem 2). Note that these conditional probabilities are decreasing in $Pr\left(m=1|S_n^N\right)$, and thus decreasing in q_n. Let μ_n be the probability that a random partner defects conditional on a player observing signal $m=1$ about the partner, and conditional on the partner observing the signal $m=0$:

$$\mu_n=\sum_{s\in S^C}Pr(s|m=1)\cdot s_0(d)+O(\epsilon_n). \tag{23}$$

Note that μ_n is decreasing in q_n. Moreover, as $q_n\nearrow\frac{1}{2\cdot k}$, we have $\mu_n(q_n)\searrow 0$, because $Pr\left(m=1|S_n^N\right)$ "explodes" as we approach the threshold of $k\cdot q=0.5$.

Next, we calculate the value of q_n that balances the payoffs of both actions when a player observes a single conflict (neglecting terms of $O(\epsilon_n)$). The LHS of the following equation represents a player's direct gain from defecting when observing a single conflict, while the RHS represents the player's indirect loss from defecting in this case, which is induced by normal partners who defect as a result of observing these defections. Note that the cost is paid only if the partner cooperated, because otherwise a future partner would observe a conflict regardless of the agent's own action.

$$\Pr(m=1)\cdot(\mu_n\cdot l+(1-\mu_n)\cdot g)=\Pr(m=1)\cdot(1-\mu_n)\cdot k\cdot q\cdot(l+1)+O(\epsilon_n)\quad\Leftrightarrow\quad q_n=\frac{\mu_n\cdot l+(1-\mu_n)\cdot g}{(1-\mu)\cdot k\cdot(l+1)}+O(\epsilon_n). \tag{24}$$

In connection with Eq. (24) it was noted that $q(\mu)$ is increasing in μ_n, and since the game is mild we have $q_n(0)=\frac{g}{k\cdot(l+1)}<\frac{1}{2\cdot k}$. This implies that there is a unique pair of values of $q_n\in\left(\frac{g}{k\cdot(l+1)},\frac{1}{2\cdot k}\right)$ and $\mu_n\in(0,1)$ that jointly solve Eqs. (23) and (24). This pair of values balances the payoff of both actions when a player observes a signal $m=1$. Note that sequences of $(q_n)_n\to q$ and $(\mu_n)_n\to\mu$ converge to the values that solve the above equations when one ignores the terms that are $O(\epsilon_n)$. The remaining arguments of part 1 are analogous to those in the final part of the proof of Theorem 2, and are omitted for brevity.

Next, we deal with Part (2), namely, the case of an acute Prisoner's Dilemma ($g>0.5\cdot(l+1)$). Assume (in order to obtain a contradiction) that the environment admits a perfect equilibrium $(S^*,\sigma^*,\theta^*\equiv c)$. That is, there exists a converging sequence of strictly positive commitment levels $\epsilon_n\to_{n\to\infty}0$, and a converging sequence of steady states $\left(S_n^N,\sigma_n,\theta_n\right)\to_{n\to\infty}(S^*,\sigma^*,\theta^*)$, such that each state $\left(S_n^N,\sigma_n,\theta_n\right)$ is a Nash equilibrium of the

perturbed environment $\left((G,k),\left(S^C,\lambda\right),\epsilon_n\right)$. By the arguments of part 1 (and the arguments of part 1(a) of Theorem 2), the average probability q_n by which a normal agent defects when observing $m=1$ in the steady state $\left(S_n^N,\sigma_n,\theta_n\right)$ (for a sufficiently small ϵ_n) should be at least equal to the minimal solution of Eq. (24): $q_n\left(\mu_n=0\right)=\frac{g}{k\cdot(l+1)}+O\left(\epsilon_n\right)$. However, if the game is acute, then this minimal solution is larger than $\frac{1}{2\cdot k}$, and Eq. (22) cannot be satisfied by $Pr\left(m=1|S_n^N\right)<<1$, which yields a contradiction.

D.9 Proof of Theorem 4 (Observing Action Profiles)

Recall that a signal $m\in M$ consists of information about the number of times in which each of the possible four action profiles have been played in the sampled k interactions. Let $u\left(m\right)$ be the number of sampled interactions in which the partner has been the sole defector, and let $d\left(m\right)$ denote the number of sampled interactions in which at least of one of the players has defected. Let s^1 and s^2 be defined as follows:

$$s^1\left(m\right)=\begin{cases}d & u\left(m\right)=1\text{ or }d\left(m\right)\geq 2\\ c & \text{otherwise}\end{cases}\qquad s^2\left(m\right)=\begin{cases}d & d\left(m\right)\geq 2\\ c & \text{otherwise.}\end{cases}$$

That is, both strategies induce agents to defect if the partner has been involved in at least two interactions in which the outcome has not been mutual cooperation. In addition, agents who follow s^1 defect also when observing the partner to be the sole defector in a single interaction.

Assume first that G_{PD} is mild (i.e., $g\leq\frac{l+1}{2}$). Fix a small probability of $0<\alpha<<\frac{1}{k}$. Let $s^\alpha\equiv\alpha$ be the strategy to defect with a probability of α regardless of the signal. In what follows, we show that there exist a converging sequence of commitment levels $\epsilon_n\to 0$ and converging sequences of steady states $\psi_n\equiv\left(\left\{s^1,s^2\right\},\left(q_n,1-q_n\right),\theta_n\right)\to\left(\left\{s^1,s^2\right\},\left(q,1-q\right),\theta^*\equiv\overrightarrow{(c,c)}\right)$, such that each steady state ψ_n is a Nash equilibrium of $\left((G,k),\left(\left\{s^\alpha\right\},1_{s^\alpha}\right),\epsilon_n\right)$.

Fix a sufficiently small $\epsilon_n<<1$. Let μ_n be the probability that the partner defects conditional on (1) the agent observing a single unilateral defection and $k-1$ mutual cooperations, i.e., $\hat{m}=\left\{(d,c),\overrightarrow{(c,c)}\right\}$ $\left(u\left(m\right)=d\left(m\right)=1\right)$, and (2) the partner observing k mutual cooperations. The parameter q_n is defined such that it balances the direct gain of defection (LHS of the equation) and its indirect loss (RHS) for a normal agent who almost always cooperates:

$$\Pr\left(\hat{m}\right)\cdot\mu_n\cdot l+(1-\mu)\cdot g=\Pr\left(\hat{m}\right)\cdot(1-\mu_n)\cdot k\cdot q_n\cdot(l+1)+O\left(\epsilon_n\right)\quad\Leftrightarrow\quad q_n=\frac{\mu_n\cdot l+(1-\mu_n)\cdot g}{(1-\mu_n)\cdot k\cdot(l+1)}+O\left(\epsilon_n\right).\tag{25}$$

The equation is the same as in the case of observation of conflicts; see Eq. (24) above. In particular, note that the indirect cost of defection when the current partner cooperates is only $O\left(\epsilon_n\right)$, because it influences only the behavior of normal future partners if they observe an additional interaction different from (c,c) in the k sampled interactions, which happens only with a probability of $O\left(\epsilon_n\right)$. Next, note that $\mu_n=\alpha\cdot Pr\left(s^\alpha|\left(d,c\right),\overrightarrow{(c,c)}\right)+O\left(\epsilon_n\right)$ because the only agents who follow s^α defect with positive probability when observing k mutual cooperations. Substituting this in (25) yields

$$q_n=\frac{g+\alpha\cdot Pr\left(s^\alpha|\left(d,c\right),\overrightarrow{(c,c)}\right)\cdot(l-g)}{\left(1-\alpha\cdot Pr\left(s^\alpha|\left(d,c\right),\overrightarrow{(c,c)}\right)\right)\cdot k\cdot(l+1)}+O\left(\epsilon_n\right)=\frac{g}{k\cdot(l+1)}+O\left(\alpha\right)+O\left(\epsilon_n\right).$$

The mildness of the game $\left(g<\frac{l+1}{2}\right)$ implies that $k\cdot q_n<0.5$.

Let p_n be the average probability with which the normal agents defect when being matched with committed

agents. When $\alpha << \frac{1}{k}$, the s^2-agents rarely ($O\left(\alpha^2\right)$) defect against the committed agents, because it is rare to observe these committed agents defecting more than once. The s^1-agents defect against the committed agents with a probability of $k \cdot q_n \cdot \alpha + O\left(\alpha^2\right) + O\left(\epsilon_n\right)$ because each rare defection of the committed agents is observed with a probability of $k \cdot q$ by s^1-agents. Since $\alpha, p_n << 1$, bilateral defections are very rare ($O\left(\alpha^2\right)$). This implies that $p_n = \alpha \cdot k \cdot q_n + O\left(\alpha^2\right) + O\left(\epsilon_n\right) < \frac{\alpha}{2}$.

Let r_n be the probability that an s^1-agent defects against a fellow s^1-agent. In each observed interaction, the s^1 partner interacts with a committed (resp., s^1, s^2) opponent with a probability of ϵ_n (resp., q_n, 1-q_n) and the partner unilaterally defects with a probability of $\alpha \cdot k \cdot q_n + O\left(\epsilon_n\right) + O\left(\alpha^2\right)$ (resp., $r_n + O\left(r_n^2\right)$, $O\left(\epsilon_n \cdot \alpha^2\right)$). This implies that r_n solves the following equation:

$$r_n = k \cdot (\alpha \cdot q \cdot \delta_n + q \cdot r_n) + O\left(\epsilon_n^2\right) \;\Rightarrow\; r_n = \frac{\alpha \cdot k \cdot q_n}{1 - k \cdot q_n} \cdot \epsilon_n + O\left(\epsilon_n^2 + \alpha^2 \cdot \epsilon_n\right) < 0.5 \cdot \alpha \cdot \epsilon_n,$$

where the latter inequality is because $k \cdot q_n < 0.5$. The above calculations show that the total frequency with which committed agents unilaterally defect ($\alpha \cdot \epsilon_n$) is higher than the total frequency with which normal agents unilaterally defect ($q_n + p_n \cdot \delta_n < \alpha \cdot \epsilon_n$). This implies that the probability that an agent is committed, conditional on his being the sole defector in an interaction, is higher than 50%, and that it is higher than this probability conditional on her being the sole cooperator. Next, note that mutual defections between a committed agent and an s^1-agent have a frequency of $O\left(\epsilon_n\right)$, while mutual defections between two committed agents (or two normal agents) are very rare ($O\left(\epsilon_n^2\right)$), which implies that the probability that the partner follows a committed strategy conditional on the player observing mutual defection is 50%+$O\left(\epsilon_n\right)$. This implies that

$$Pr\left(s^{\alpha} | \left(d,c\right), \overrightarrow{(c,c)}\right) > \max\left(Pr\left(s^{\alpha} | \left(d,d\right), \overrightarrow{(c,c)}\right), Pr\left(s^{\alpha} | \left(c,d\right), \overrightarrow{(c,c)}\right)\right),$$

and thus while both actions are best replies after the player observes the signal $\left(\left(d,c\right), \overrightarrow{(c,c)}\right)$, only cooperation is a best reply after the player observes $\left(\left(d,d\right), \overrightarrow{(c,c)}\right)$ and $\left(\left(c,d\right), \overrightarrow{(c,c)}\right)$. Next note that conditional on a player observing a signal with at most $k-2$ mutual cooperations, the partner is most likely to be committed (because normal agents have two outcomes different from mutual cooperation with a probability of only $O\left(\epsilon_n^2\right)$). This implies that the normal agents play the unique best reply after any signal other than $\left(\left(d,c\right), \overrightarrow{(c,c)}\right)$, and thus any deviator who behaves differently in these cases will be outperformed.

Let χ_n be the probability that a random partner defects conditional on both the agent and the partner observing signal $\left(\left(d,c\right), \overrightarrow{(c,c)}\right)$. The definitions of strategies s^{α}, s^1, and s^2 immediately imply that $\chi_n > \mu_n$, and analogous arguments to those presented at the end of the proof of Theorem 2 show that deviators who defect with a probability strictly between zero and one after observing $\left(\left(d,c\right), \overrightarrow{(c,c)}\right)$ are outperformed (because an agent's payoff is a strictly convex function of the agent's defection probability when observing signal $\left(\left(d,c\right), \overrightarrow{(c,c)}\right)$).

Next assume that the G_{PD} is acute. We have to show that cooperation is not a perfect equilibrium action. Assume to the contrary that $(S^*, \sigma^*, \theta^* \equiv 0)$ is a perfect equilibrium with respect to distribution of commitments $\left(\mathcal{S}^{\mathcal{C}}, \lambda\right)$. Let $\psi_n = \left(S_n^N, \sigma_n, \theta_n\right) \rightarrow (S^*, \sigma^*, 0)$ be a converging sequence of Nash equilibria in the converging sequence of perturbed environments $\left(\left(G_{PD}, k\right), \left(\mathcal{S}^{\mathcal{C}}, \lambda\right), \epsilon_n\right)$. Analogous arguments to the proof of part 1(a) of Theorem 2 show that any perfect equilibrium that implements full cooperation $(S^*, \sigma^*, \theta^* \equiv 0)$ must satisfy (1) $s_{\overrightarrow{(c,c)}} = c$ for each $s \in S^*$, (2) if $d\left(m\right) \geq 2$ then $s_m = d$ for each $s \in S^*$, and (3) there are $s, s' \in S^*$ such that $s_{\left\{(d,c), \overrightarrow{(c,c)}\right\}}\left(d\right) > 0$ and $s_{\left\{(d,c), \overrightarrow{(c,c)}\right\}}\left(d\right) < 1$.

Let $0 < q_n < 1$ be the average probability according to which a normal agent defects when she observes $\left\{\left(d,c\right), \overrightarrow{(c,c)}\right\}$. By analogous arguments to those presented above (see Eq. (25)), q_n is an increasing function

of μ_n, and $q_n\left(\mu_n=0\right)=\frac{g}{k\cdot(l+1)}$. The acuteness of the game implies that $k\cdot q_n>\frac{g}{(l+1)}>\frac{1}{2}$.

Let $s_\beta\in S^C$ be a committed strategy that induces an agent who follows it (called an s_β-*agent*) to defect with a probability of $\beta>0$ when he observes $\left(\overrightarrow{(c,c)}\right)$. In what follows, we show that the presence of strategy s_β induces the normal agents to unilaterally defect more often than s_β-agents. Let p_n be the average probability that normal agents defect against s^α-agents in state ψ_n. This probability p_n must solve the following inequality:

$$\begin{aligned}1-p_n \quad\geq\quad & \left((1-\beta)\cdot(1-p_n)\right)^k+k\cdot\left((1-\beta)\cdot(1-p_n)\right)^{k-1}\cdot\left(1-(1-\beta)\cdot(1-p_n)\right)\\ & +(1-q_n)\cdot k\cdot\left((1-\beta)\cdot(1-p_n)\right)^{k-1}\cdot\beta\cdot(1-p_n)+O\left(\epsilon_n\right).\end{aligned}\tag{26}$$

The LHS of Eq. (26) is the average probability that normal agents cooperate against s_β-agents (recall that normal agents always defect when they observe at most $k-2$ mutual cooperations). The normal agents cooperate with probability one (resp., at most one, q_n) if they observe $\left(\overrightarrow{(c,c)}\right)$ (resp., $\left((d,d),\overrightarrow{(c,c)}\right)$ or $\left((c,d),\overrightarrow{(c,c)}\right)$, $\left((d,c),\overrightarrow{(c,c)}\right)$), which happens with a probability of $\left((1-\beta)\cdot(1-p_n)\right)^k$ (resp., $k\cdot\left((1-\beta)\cdot(1-p_n)\right)^{k-1}\cdot\left(1-(1-\beta)\cdot(1-p_n)\right)$, $k\cdot\left((1-\alpha)\cdot(1-p_n)\right)^{k-1}\cdot\alpha\cdot(1-p)$).

Direct numerical analysis of Eq. (26) shows that the minimal p_n that solves this inequality (given that $q_n>\frac{1}{2\cdot k}$) is greater than $\frac{\beta}{2-\beta}$ for any $\beta\in(0,1)$. The total frequency of interactions in which the s_β-agents unilaterally defect is $\beta\cdot(1-p_n)\cdot\epsilon_n\cdot\lambda\left(s_\beta\right)+O\left(\epsilon_n^2\right)$. The total frequency of interactions in which normal agents unilaterally defect against the s_β-agents is $p_n\cdot(1-\beta)\cdot\epsilon_n\cdot\lambda\left(s_\beta\right)+O\left(\epsilon_n^2\right)$. Eq. (25) shows that these unilateral defections against s_β-agents induce the normal agents to unilaterally defect among themselves with a total frequency of $\frac{p_n\cdot(1-\beta)\cdot\epsilon_n\cdot\lambda(s_\beta)}{1-k\cdot q_n}+O\left(\epsilon_n^2\right)>p_n\cdot(1-\beta)\cdot\epsilon_n\cdot\lambda\left(s_\beta\right)$. Finally, note that $p_n>\frac{\beta}{2-\beta}\Leftrightarrow 2\cdot p_n\cdot(1-\beta)>\beta\cdot(1-p_n)$ implies that normal agents unilaterally defect (as the indirect result of the presence of the s_β-agents) more often than s_β-agents.

Next, observe that bilateral defections are most likely to occur in interactions between normal and committed agents. This is because the probability that both normal agents defect against each other is only $O\left(\epsilon_n^2\right)$. Thus, when a player observes bilateral defection the partner is more likely to be a committed agent than when the player observes a unilateral defection by the partner. This implies that all the normal agents defect with probability one when they observe $\left((d,d),\overrightarrow{(c,c)}\right)$ because in this case defection is the unique best reply.

Let w_n be the (average) probability that normal agents defect when they observe $\left((c,d),\overrightarrow{(c,c)}\right)$. If $w_n<0.5$, then cooperation is the unique best reply for a normal agent who faces a partner who is likely to defect (e.g., when the normal agent observes fewer than $k-1$ mutual cooperations), and so we get a contradiction. This is because defecting against a defector yields a direct gain of l and an indirect loss of at least $0.5\cdot k\cdot(l+1)\geq l+1>l$ (because this bilateral defection will be observed on average k times, and in at least half of these cases it will induce the partner to defect, whereas if the agent were cooperating, then he would have induced the partner to cooperate).

Thus, $w_n\geq 0.5\Rightarrow k\cdot w_n>1$. However, in this case, an analogous argument to the one at the end of the proof of Theorem 3 implies that an arbitrarily small group of mutants who defect with small probability will cause the incumbents to unilaterally defect with high probability, and thus no focal post-entry population exists, which contradicts the assumption that cooperation is neutrally stable.

D.10 Proof of Theorem 5 (Observing Actions against Cooperation)

The construction of the distribution of commitments $\left(\left\{s^\alpha\right\},1_{s^\alpha}\right)$ and of the perfect equilibrium $\left(\left\{s^1,s^2\right\},(q,1-q),\theta\equiv\overrightarrow{(c,c)}\right)$ and most of the arguments are the same as in the proof of Theorem 4, and are

omitted for brevity. Fix ϵ_n sufficiently small. By the same arguments as in the proof of Theorem of 3, the value of q_n that balances the payoffs of s^1 and s^2 satisfy $k \cdot q_n < 1$ for any underlying Prisoner's Dilemma.

Recall that p_n, the average probability with which the normal agents defect when being matched with committed agents, satisfies $p_n = \alpha \cdot k \cdot q_n + O\left(\alpha^2\right) + O\left(\epsilon_n\right) < \alpha$. This implies that the probability that an agent is committed, conditional on her being the sole defector in an interaction, is higher than 50%, conditional on her being the sole cooperator. Next, observe that $\alpha << 1$ implies that the probability $Pr\left((d,d)\right) = O\left(p_n \cdot \alpha^2\right) \cdot O\left(\epsilon_n\right) << Pr\left((c,d)\right) = O\left(p_n \cdot \epsilon_n \cdot \alpha\right)$, which implies that conditional on an agent observing the signal $\left\{(*,d), \left(\overrightarrow{(c,c)}\right)\right\}$, it is most likely that the partner has cooperated rather than defected in the interaction in which $(*,d)$ has been observed. This implies that $\Pr\left(s^\alpha | \left\{(*,d), \left(\overrightarrow{(c,c)}\right)\right\}\right) < \Pr\left(s^\alpha | \left\{(d,c), \left(\overrightarrow{(c,c)}\right)\right\}\right)$, and given the value of q_n for which both actions are best replies conditional on observing signal $\left\{(d,c), \left(\overrightarrow{(c,c)}\right)\right\}$, cooperation is the unique best reply when observing either $\left\{(*,d), \left(\overrightarrow{(c,c)}\right)\right\}$ or $\left\{\left(\overrightarrow{(c,c)}\right)\right\}$, while defection is the unique best reply when observing at most $k-2$ mutual cooperations. This implies that $\left(\left\{s^1, s^2\right\}, (q, 1-q), \theta \equiv 0\right)$ is a perfect equilibrium (where q is the limit of q_n when ϵ_n converges to zero.

D.11 Proof of Theorem 6 (Repeated Game)

Part 1: Assume that $g > l$ (i.e., a defensive game). Assume to the contrary that there exist a sequence of Nash equilibria of perturbed environments that converge to a perfect equilibrium that induces full cooperation. The fact that the perfect equilibrium induces full cooperation implies that in any sufficiently close Nash equilibrium (i.e., for a sufficiently large n):

1. normal agents cooperate with high probability when observing $(c, ..., c)$;
2. most of the time when an agent is matched with a normal partner, the agent observes the signal $(c, ..., c)$;
3. when a normal agent observes the signal $(c, ..., c)$ the partner is most likely normal and he is going to cooperate with a probability close to one;
4. when an agent observes the signal $(d, .., d)$ the partner is most likely committed and is going to defect with positive probability.

In order for these facts to be be consistent with equilibrium it must be the case that cooperation is a best reply against a partner who is most likely to cooperate in the current match, i.e., the direct gain from defecting, which is very close to g, has to be lower than the future indirect loss, which is independent of the partner's action. The inequality $g > l$ then implies that cooperation is the unique best reply against a partner who is going to cooperate with an expected probability that is not close to 1 (because the direct gain from defecting is a mixed average of l and g, which is less than g). This, in turn, implies that all normal agents cooperate with a probability of one when they observe the signal $(d, .., d)$ (because, given such a signal, the partner is most likely to be a committed agent and to defect with a positive probability in the current match). Hence, a deviator who always defects outperforms the incumbent, since she induces normal agents to cooperate against her, and obtains the high payoff of $1 + g$ in most rounds of the repeated game.

Part 2: Assume that $g \leq l$ (i.e., a defensive game). Let $\gamma = \frac{l}{\delta \cdot (l+1)} \in (0,1)$. Let $0 < \underline{\alpha} < \overline{\alpha} < 1$ be two probabilities satisfying the condition that the ratio $\overline{\alpha}/\underline{\alpha}$ is sufficiently large (as further specified below). Consider a homogeneous group of committed agents who

1. defect with probability $\overline{\alpha}$ if they either (1) defected in the last round, or (2) defected at least twice in the last $k-1$ rounds; and

2. defect with probability $\underline{\alpha}$ otherwise.

Consider the perturbed environment $\left((G,k,\delta),\left(S^C,1_{S^C}\right),\epsilon_n\right)$, for a sufficiently small $\epsilon>0$. Consider a homogeneous population of normal agents who play according to the following strategy s_N:

1. cooperate if the agent defected in any of the last $\min(t,k-1)$ rounds;
2. otherwise (i.e., the agent cooperated in all of the last $\min(t,k-1)$ rounds):
   (a) cooperate if the partner has never defected in the last $\min(t,k-1)$ rounds;
   (b) defect if the partner defected at least twice in the last $\min(t,k-1)$ rounds;
   (c) cooperate if the partner defected only once in the last $\min(t,k-1)$ rounds and did not defect in the last round; and
   (d) defect with probability q_t if the partner defected only in the last round, where t is the current round, and the sequence $(q_t)_{t\geq 1}$ is defined recursively below.

Let $q_1=\gamma$. The value of each q_t for $t\geq 2$ is determined such that a normal agent is indifferent between defecting and cooperating in round $t-1$ conditional on the events that (1) the agent did not defect in any of the previous $k-1$ rounds, and (2) the agent observes the signal $(c,...,c,d)$ (i.e., the partner defected in the last round and cooperated in all of the previous observed interactions). Here we are relying on the one-deviation principle; in the next period the agent will have a track record $(c,c,c,...c,d)$, which means that the agent should cooperate.

The gain from defecting in round $t-1$ is equal to $l\cdot\mu_{t-1}+g\cdot(1-\mu_{t-1})$, where μ_{t-1} is the probability that a random partner defects conditional on the union of the two events above. Such a defection induces an expected loss of $\delta\cdot(l+1)\cdot q_t+O(\epsilon)$ for the agent in the next round (with a probability of $(1-\epsilon)$ the partner in the next round is normal, and in this case he will defect with probability q_t instead of cooperating, which will induce a loss of $\delta(l+1)$ for the agent. In the round after that the agent will have a track record $(c,c,c,...c,d,c)$ which means that the agent should cooperate again. The partner, if normal, will cooperate for sure with the agent. Thus, an agent is indifferent between the two actions in round $t-1$ when observing $(c,...,c,d)$ iff

$$l\cdot\mu_{t-1}+g\cdot(1-\mu_{t-1})=\delta\cdot(l+1)\cdot q_t+O(\epsilon)\Leftrightarrow q_t=\frac{l\cdot\mu_{t-1}+g\cdot(1-\mu_{t-1})}{\delta\cdot(l+1)}+O(\epsilon).$$

Observe that the q_t's have a uniform bound strictly below one, i.e., $\forall t\in\mathbb{N}$, $0<q_t<\gamma=\frac{l}{\delta\cdot(l+1)}<1$. Let $(\beta_t)_{t\in\mathbb{N}}$ be the average probability with which normal agents defect in round t. Observe that $\beta_1=0$, and that β_t can be bounded as follows for any $t\in\mathbb{N}$:

$$\beta_t\leq q_t\cdot\beta_{t-1}+O(\epsilon)<\gamma\cdot\beta_{t-1}+O(\epsilon).$$

This implies that β_t is bounded from above by a converging geometric sequence, and, thus, $\beta_t<\frac{O(\epsilon)}{1-\gamma}$ for each t. This implies that the population state $(s_N,1_{s_N})$ induces full cooperation in the limit $\epsilon\to 0$.

Let $Pr\left(S^C|\,(c,...,c,d),t\right)$ be the probability that the partner is committed conditional on the agent observing signal $(c,...,c,d)$ in round t. Let $Pr\left((c,...,c,d),t|S^C\right)$ $(Pr\left((c,...,c,d),t|s_N\right))$ be the probability that an agent observes the signal $(c,...,c,d)$ in round t conditional on the partner being committed (normal). Observe that $Pr\left((c,...,c,d),t|S^C\right)>\underline{\alpha}^k$ (because a committed agent plays each pure action with a probability of at least $\underline{\alpha}$ in each round), and that $Pr\left((c,...,c,d),t|S^C\right)<sup_t\beta_t<\frac{O(\epsilon)}{1-\gamma}$ (because the average probability in which a normal agent defects is at most $sup_t\beta_t$. By using Bayes' rule we can give a uniform minimal bound to

$Pr\left(S^C|\left(c,...,c,d\right),t\right)$ as follows:

$$\frac{\epsilon\cdot\underline{\alpha}^k}{\epsilon\cdot\underline{\alpha}^k+\frac{O(\epsilon)}{1-\gamma}}<Pr\left(S^C|\left(c,...,c,d\right),t\right)<1.$$

We assume that the ratio $\overline{\alpha}/\underline{\alpha}$ is sufficiently large such that $\overline{\alpha}\cdot Pr\left(S^C|\left(c,...,c,d\right),t\right)>\underline{\alpha}$ in each round t. Recall the definition from above and then observe that $\mu_{t-1}=Pr\left(S^C|\left(c,...,c,d\right),t\right)\cdot\bar{\alpha}\in(\underline{\alpha},\bar{\alpha})$ for each round t. Recall that the probabilities $(q_t)_{t\in\mathbb{N}}$ have been defined such that each normal agent is indifferent between the two actions when (1) she observes the signal $(c,...,c,d)$, and (2) she did not defect in any of the previous $k-1$ rounds. Next we show that the normal agents have strict preferences in all other cases. Specifically, the fact that $g\leq l$ (resp., $g<l$) implies that each normal agent:

1. strictly prefers to cooperate if she defected in any of the last $k-1$ rounds, because otherwise future normal incumbents will defect for sure in at least one additional future round (inducing an indirect loss of at least $\delta^k\cdot(l+1)>l>g$, which is larger than the agent's direct gain of defection);

2. weakly (resp., strictly) prefers to cooperate if she observes the signal $(c,...,c)$; in this case, the partner is most likely a normal agent who is going to cooperate, and the direct gain from defecting (g) is outweighed by the larger indirect loss in the next round $(\delta\cdot(l+1)\cdot q_t+O(\epsilon)>g)$.

3. weakly (resp., strictly) prefers to defect if (1) the partner defected at least twice in the last k rounds, and (2) the agent did not defect in any of the last $k-1$ rounds; in this case the partner is most likely to be a committed agent and to defect with a high probability of $\overline{\alpha}>\mu_{t-1}$ in each round t and, thus, defection is the agent's unique best reply.

4. weakly (resp., strictly) prefers to cooperate if the partner defected only once in the last k rounds, and this defection did not happen in the last round; in this case the probability that the partner is going to defect in the current match is at most $\underline{\alpha}<\mu_{t-1}$ for each round t and, thus, cooperation is the agent's unique best reply.

This implies that the population state $(s_N,1_{s_N})$ is indeed a Nash equilibrium of the perturbed environment for a sufficiently small ϵ.